TESIS-SS14 2501

PEMODELAN REGRESI ZERO INFLATED NEGATIVE BINOMIAL (ZINB) PADA KASUS TETANUS NEONATORUM DI PROVINSI JAWA TIMUR

CINDY CAHYANING ASTUTI
NRP. 1313201008

DOSEN PEMBIMBING
Dr. Ismaini Zain, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2015

THESIS-SS14 2501

MODELLING ZERO INFLATED NEGATIVE BINOMIAL (ZINB) WITH AN APPLICATION ON THE CASE OF TETANUS NEONATORUM IN EAST JAVA

CINDY CAHYANING ASTUTI
NRP. 1313201008

ADVISOR
Dr. Ismaini Zain, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2015

PEMODELAN REGRESI ZERO INFLATED NEGATIVE BINOMIAL (ZINB) PADA KASUS TETANUS NEONATORUM DI PROVINSI JAWA TIMUR

Tesis ini disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si)
di
Institut Teknologi Sepuluh Nopember

Oleh :

CINDY CAHYANING ASTUTI
NRP 1313201008

Tanggal Ujian : 25 Mei 2015
Periode Wisuda : September 2015

Disetujui Oleh :

1. **Dr. Ismaini Zain, M.Si** **(Pembimbing)**
   NIP 19600525 198803 2 001

Ratnasari -

2. **Dr. Vita Ratnasari, M.Si** **(Penguji)**
   NIP 19700910 199702 2 001

3. **Dr. Santi Puteri Rahayu, M.Si** **(Penguji)**
   NIP. 19750115 199903 2 003

Direktur Program Pascasarjana

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
PASCASARJANA

Prof. Dr. Ir. Adi Soeprijanto, M.T
NIP 19640405 199002 1 001

PEMODELAN REGRESI ZERO INFLATED NEGATIVE BINOMIAL (ZINB) PADA KASUS TETANUS NEONATORUM DI PROVINSI JAWA TIMUR

Nama : Cindy Cahyaning Astuti
NRP : 1313 201 008
Pembimbing : Dr. Ismaini Zain, M.Si

ABSTRAK

Analisis regresi digunakan untuk mengetahui hubungan antara satu atau beberapa variabel respon (Y) dengan satu atau beberapa variabel prediktor (X). Model regresi yang digunakan untuk menjelaskan hubungan antara variabel prediktor dan variabel respon yang memiliki sebaran *Poisson* adalah model regresi *Poisson*. Namun, pada model regresi *Poisson* terdapat asumsi ragam harus sama dengan rata-rata (*equidispersion*), sehingga model ini tidak tepat digunakan pada data yang mengalami *overdispersion* (ragam lebih besar dari rata-rata). Regresi *Poisson* adalah model umum yang digunakan untuk menganalisis *count data* (data hitung). Pada jenis *count data* (data hitung) sering dijumpai amatan yang bernilai nol dengan proporsi nilai nol yang besar pada variabel respon (*zero inflation*). Regresi *Poisson* dapat digunakan untuk menganalisis data hitung namun masih belum dapat mengatasi masalah nilai nol berlebih pada variabel respon (*zero inflation*). Alternatif model yang lebih sesuai untuk data yang mengalami *overdispersion* dan dapat mengatasi masalah nilai nol berlebih pada variabel respon (*zero inflation*) adalah model regresi *Zero Inflated Negative Binomial* (ZINB). Model regresi *Zero Inflated Negative Binomial* (ZINB) diaplikasikan pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur. Tujuan penelitian ini adalah mengkaji bentuk *likelihood* dan mengkaji estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) serta mengaplikasikan model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur. Hasil pengujian parameter model regresi ZINB menunjukkan bahwa variabel prediktor yang berpengaruh signifikan secara parsial pada model *negative binomial* adalah persentase kunjungan ibu hamil dan persentase ibu bersalin ditolong tenaga kesehatan, sedangkan variabel prediktor yang berpengaruh signifikan secara parsial pada model *zero inflation* adalah persentase kunjungan neonatus.

Kata Kunci : *Overdispersion, Tetanus Neonatorum, Zero Inflation*, *Zero Inflated Negative Binomial* (ZINB).

(Halaman ini sengaja dikosongkan)

MODELLING ZERO INFLATED NEGATIVE BINOMIAL WITH AN APPLICATION ON THE CASE OF TETANUS NEONATORUM IN EAST JAVA

Name : Cindy Cahyaning Astuti
Student Id. Number : 1313 201 008
Advisor : Dr. Ismaini Zain, M.Si

ABSTRACT

Regression analysis is used to determine relationship between one or several response variable (Y) with one or several predictor variables (X). Regression model between predictor variables and the Poisson distributed response variable is called Poisson Regression Model. Since, Poisson Regression requires an equality between mean and variance, it is not appropriate to apply this model on overdispersion (variance is higher than mean). Poisson regression model is commonly used to analyze the count data. On the count data type, it is often to encounteredd some observations that have zero value with large proportion of zero value on the response variable (zero Inflation). Poisson regression can be used to analyze count data but it has not been able to solve problem of excess zero value on the response variable. An alternative model which is more suitable for overdispersion data and can solve the problem of excess zero value on the response variable is Zero Inflated Negative Binomial (ZINB). In this research, ZINB is applied on the case of Tetanus Neonatorum in East Java. The aim of this research is to examine the likelihood function and to form an algorithm to estimate the parameter of ZINB and also applying ZINB model in the case of Tetanus Neonatorum in East Java. Maximum Likelihood Estimation (MLE) method is used to estimate the parameter on ZINB and the likelihood function is maximized using Expectation Maximization (EM) algorithm. Test results of ZINB regression model showed that the predictor variable have a partial significant effect at negative binomial model is the percentage of pregnant women visits and the percentage of maternal health personnel assisted, while the predictor variables that have a partial significant effect at zero inflation model is the percentage of neonatus visits.

Keyword: Overdispersion, Tetanus Neonatorum, Zero Inflation, Zero Inflated Negative Binomial (ZINB).

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga tesis ini dapat diselesaikan tepat pada waktunya. Tesis ini disusun untuk memenuhi salah satu persyaratan dalam rangka menyelesaikan pendidikan pada Program Magister Jurusan Statistika FMIPA ITS. Tesis ini berjudul: **"Pemodelan Regresi Zero Inflated Negative Binomial (ZINB) Pada Kasus Tetanus Neonatorum Di Provinsi Jawa Timur"**.

Dalam penyusunan tesis ini, penulis banyak memperoleh bimbingan dan petunjuk, serta bantuan dan dukungan dari berbagai pihak baik dari institusi maupun luar institusi. Melalui kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Ibu Dr. Ismaini Zain, M.Si selaku dosen pembimbing,yang telah banyak meluangkan waktunya untuk memberikan arahan dalam menyelesaikan tesis ini serta nasehat untuk menjadi lebih baik.
2. Ibu Dr. Vita Ratnasari, M.Si dan ibu Dr. Santi Puteri Rahayu, M.Si selaku dosen penguji yang telah memberikan saran serta perbaikan dalam tesis ini.
3. Bapak Dr. Mashuri, M.T selaku ketua Jurusan Statistika FMIPA ITS.
4. Bapak Dr. Suhartono, M.Sc. selaku ketua Program Studi Pascasarjana Jurusan Statistika ITS.
5. Ibu Dra. Destri Susilaningrum, M.Si selaku Dosen Wali.
6. Bapak dan Ibu dosen pengajar Jurusan Statistika ITS, terima kasih atas ilmu yang telah diajarkan.
7. Bapak-bapak dan Ibu-ibu Pegawai Jurusan Statistika ITS yang telah banyak membantu penulis selama masa perkuliahan.
8. Kedua Orang tua tercinta Bapak H. Suranto dan Ibu Hj. Mas'ulah, adik M. Kresno Gumelar, serta seluruh keluarga besar yang selalu mendoakan dan memberikan dukungan.

9. Sahabat-sahabat Statistika Universitas Brawijaya ’09 hadi, pika, witra, asri, delbra, erwin, hanafi, benu, danang terimakasih atas canda tawa, semangat dan kebersamaannya.
10. Sahabat-sahabat tercinta mbak luthfah, mbak vita, mbak fenda, mbak ami, mbak fifi, mirah, mbak nanik, mbak ina, mbak farida, safitri, mas untung, mas jihad, mas zul, mas ikbal terimakasih banyak sudah banyak direpotkan dalam segala hal dan atas semua doa serta dukungannya.
11. Sahabat seperjuangan bimbingan mbak irun dan evellin, terimakasih atas semangat dan dukungannya.
12. Semua pihak yang tidak sempat disebutkan satu-persatu atas doa dandukungan yang telah diberikan kepada penulis selama ini.

Akhirnya, do’a dan harapan selalu dipanjatkan kepada Allah SWT agar ilmu yang telah diperoleh menjadi barokah dan bermanfaat bagi sesama sertadapat menjadi sarana meraih ridho-Nya. Aamiin Ya Robbal ‘Alamin.

Surabaya, Juni 2015

Cindy Cahyaning Astuti

DAFTAR ISI

DAFTAR TABEL

Halaman

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS

CINDY CAHYANING ASTUTI lahir di Desa Pilang Kabupaten Sidoarjo, Jawa Timur pada tanggal 14 Juli 1991. Anak pertama dari dua bersaudara dari pasangan Bapak H. Suranto dan Ibu Hj. Mas'ulah ini menyelesaikan pendidikan sekolah dasar di SDN Pilang 1 tahun 2003 kemudian melanjutkan di SMP Negeri 4 Sidoarjo dan selesai pada tahun 2006. Pendidikan selanjutnya di SMA Negeri 3 Sidoarjo hingga lulus pada tahun 2009. Pendidikan Tinggi dimulai pada tahun 2009 di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Brawijaya, Program Studi Statistika melalui jalur PMDK dan lulus pada tahun 2013. Pada tahun yang sama, tahun 2013 melanjutkan pendidikan s2 melalui program Beasiswa Pendidikan Pascasarjana Dalam Negeri tahun 2013-2015 oleh Direktorat Jenderal Pendidikan Tinggi (Ditjen Dikti) di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Surabaya, Juni 2015

(cindycahyaning.a@gmail.com)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi digunakan untuk mengetahui hubungan antara satu atau beberapa variabel respon (Y) dengan satu atau beberapa variabel prediktor (X). Pada model linier klasik terdapat asumsi variabel respon mengikuti sebaran normal, namun pada kenyataan sering ditemukan kondisi variabel respon tidak mengikuti sebaran normal. Menurut Agresti (2002), untuk mengatasi hal tersebut terdapat pengembangan dalam model linier klasik yaitu *Generalized Linear Model* (GLM). GLM mengasumsikan variabel respon mengikuti sebaran keluarga eksponensial, yang memiliki sifat lebih umum. Beberapa sebaran yang termasuk dalam sebaran keluarga eksponensial adalah sebaran *Normal, Poisson, Binomial, Exponensial* dan *Gamma.*

Pada berbagai penelitian, sering dijumpai data dengan variabel respon yang mengikuti sebaran *Poisson*, analisis regresi yang digunakan untuk data seperti ini adalah analisis regresi *Poisson*. Regresi *Poisson* adalah model umum yang digunakan untuk menganalisis *count data* (data hitung). Dalam regresi *Poisson* terdapat asumsi Y~ *Poisson* (μ), hal ini berarti variabel respon diasumsikan menyebar *Poisson*. Asumsi penting pada analisis regresi *Poisson* adalah ragam harus sama dengan rata-rata, kondisi ini disebut *equidispersion*. Menurut Famoye dan Singh (2006), pada jenis *count data* (data hitung) sering dijumpai kondisi terdapat nilai nol yang lebih dari 50 persen pada variabel respon (*zero inflation*). Proporsi data yang memiliki nilai nol berlebihan ini dapat berakibat pada ketepatan dari inferensia. Regresi *Poisson* dapat digunakan untuk menganalisis data hitung namun masih belum dapat mengatasi masalah nilai nol berlebihan pada variabel respon (*zero inflation*).

Menurut Lambert (1992), jika pada suatu pemodelan *count data* (data hitung) banyak terdapat amatan yang bernilai nol pada variabel respon (*zero inflation*) maka dapat diatasi dengan menggunakan model regresi *Zero Inflated Poisson* (ZIP). Namun apabila terdapat data dengan banyak amatan yang bernilai nol dan terjadi *overdispersion* maka model regresi *Zero Inflate*d *Poisson* (ZIP) sudah tidak tepat lagi digunakan. Kondisi *overdispersion* dapat didefinisikan sebagai kondisi dalam sebaran *Poisson* dimana ragam lebih besar dari rata-rata. Menurut Hinde dan Demetrio (2007), *overdispersion* pada regresi *Poisson* dapat mengakibatkan *standard error* dari estimasi parameter regresi yang dihasilkan memiliki kecenderungan untuk menjadi lebih rendah dari seharusnya, sehingga menghasilkan kesimpulan yang tidak sesuai dengan data. Jika pada suatu pemodelan *count data* (data hitung) banyak terdapat amatan yang bernilai nol pada variabel respon (*zero inflation*) dan terjadi *overdispersion* maka model yang dapat digunakan adalah model regresi *Zero Inflated Generalized Poisson* (Famoye & Singh, 2006).

Pada perkembangannya terdapat alternatif lain untuk memodelkan kasus dengan banyak amatan yang bernilai nol dan terjadi *overdispersion* selain menggunakan model regresi *Zero Inflated Generalized Poisson* (ZIGP), model tersebut adalah regresi *Zero Inflated Negative Binomial* (ZINB). Menurut Hilbe (2011), model regresi *Zero Inflated Negative Binomial* merupakan model yang dibentuk dari sebaran campuran *Poisson Gamma*. Model regresi *Zero Inflated Negative Binomial* (ZINB) dapat digunakan sebagai alternatif lain dalam memodelkan kasus dengan banyak amatan yang bernilai nol dan terjadi *overdispersion* karena model ini tidak mensyaratkan ragam harus sama dengan rata-rata, selain itu model regresi *Zero Inflated Negative Binomial* (ZINB) juga memiliki parameter dispersi yang berguna untuk menggambarkan variasi dari data, yang biasa dinotasikan dengan $\boldsymbol{\kappa}$ (kappa).

Pemodelan regresi *Zero Inflated Poisson* (ZIP) dan regresi *Zero Inflated Generalized Poisson* (ZIGP) telah banyak dilakukan yaitu oleh Lambert (1992), Famoye dan Singh (2006), Lestari (2009) dan Lestari (2014). Lambert (1992), menggunakan model regresi *Zero Inflated Poisson* (ZIP) dengan mengaplikasikan model tersebut pada data yang dikumpulkan dari sebuah studi *Quality Control*. Famoye dan Singh (2006), menggunakan model regresi *Zero Inflated Generalized Poisson* (ZIGP) untuk menganalisis kasus kekerasan dalam rumah tangga. Lestari (2009), menggunakan regresi *Zero Inflated Poisson* (ZIP) untuk memodelkan data pekerja seks komersial dan Lestari (2014) menggunakan regresi *Zero Inflated*

Generalized Poisson (ZIGP) untuk memodelkan data penderita *Tetanus Neonatorum*.

Sepengetahuan penulis, belum ada kajian secara mendalam tentang model regresi *Zero Inflated Negative Binomial* (ZINB). Oleh karena itu pada penelitian ini akan dilakukan pemodelan regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur. Pada penelitian Lestari (2014), terdapat empat variabel prediktor yang digunakan untuk memodelkan data penderita *Tetanus Neonatorum* menggunakan regresi *Zero Inflated Generalized Poisson* (ZIGP) yaitu persentase ibu bersalin ditolong tenaga kesehatan, persentase ibu bersalin ditolong dukun, persentase kunjungan ibu hamil K4 dan persentase kunjungan neonatus. Menurut Saifuddin *et al*. (2006), terdapat faktor penting yang secara efektif bisa mencegah terjadinya penyakit *Tetanus Neonatorum* yaitu Imunisasi Toksoid (TT) pada ibu hamil. Berdasarkan informasi tersebut maka pada penelitian ini akan ditambahkan variabel prediktor sebagai faktor pencegahan terjadinya penyakit *Tetanus Neonatorum* yaitu Imunisasi Toksoid (TT) pada ibu hamil. Pada penelitian pendahuluan, data penderita *Tetanus Neonatorum* di Provinsi Jawa Timur adalah data yang memiliki sebaran *Poisson* dan terjadi *overdispersion* serta memiliki proporsi nilai nol yang besar yaitu 73,7 persen, oleh karena itu untuk memodelkan data digunakan regresi *Zero Inflated Negative Binomial* (ZINB).

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian dirumuskan permasalahan sebagai berikut.

1. Bagaimana bentuk *likelihood* dan bagaimana kajian estimasi parameter pada model regresi *Zero Inflated Negative Binomial* (ZINB)?
2. Bagaimana aplikasi model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut.

1. Mengkaji bentuk *likelihood* dan mengkaji estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB).
2. Mengaplikasikan model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur**.**

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini adalah sebagai berikut.

1. Mengembangkan wawasan pengetahuan yang berkaitan dengan bentuk *likelihood* dan kajian estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) serta aplikasinya pada studi kasus tentang penyakit *Tetanus Neonatorum* di Provinsi Jawa Timur.
2. Memberikan informasi mengenai faktor-faktor yang berpengaruh terhadap kasus penyakit *Tetanus Neonatorum* di Provinsi Jawa Timur dengan menggunakan model regresi *Zero Inflated Negative Binomial* (ZINB).

1.5 Batasan Masalah

Pada penelitian ini, estimasi parameter pada model regresi *Zero Inflated Negative Binomial* (ZINB) dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan untuk memaksimalkan fungsi *likelihood* digunakan algoritma EM (*Expectation Maximization*).

BAB 2
TINJAUAN PUSTAKA

Pada bab tinjauan pustaka ini dijelaskan beberapa teori terkait yang mendukung penyelesaian masalah dalam penelitian ini. Beberapa hal yang akan dibahas pada bab ini adalah : *Generalized Linier Model* (GLM), Sebaran *Poisson*, Regresi *Poisson*, Multikolinieritas, Regresi *Negative Binomial*, Regresi *Zero Inflated Negative Binomial* dan kajian non statistik meliputi bahasan tentang penyakit *Tetanus Neonatorum*.

2.1 *Generalized Linear Model* (GLM)

Generalized Linear Model (GLM) merupakan perluasan model regresi umum untuk variabel respon mengikuti sebaran keluarga eksponensial. Menurut Agresti (2002), terdapat tiga komponen dalam GLM yaitu:

1. *random component* (komponen acak) yaitu komponen yang ditunjukkan dengan variabel respon Y yang bersifat independen dan mengikuti sebaran keluarga eksponensial;
2. *systematic component* (komponen sistematik) berhubungan dengan variabel prediktor yang digunakan. Komponen sistematik yaitu vector $\boldsymbol{\eta}$ yang terdiri dari $[\eta_1,\eta_2,\ldots,\eta_n]^T$ yang memiliki bentuk umum dari $\boldsymbol{\eta} = \boldsymbol{X\beta}$ di mana $\boldsymbol{X}$ merupakan suatu matriks dengan elemen yang terdiri variabel prediktor, sedangkan $\boldsymbol{\beta}$ merupakan bentuk vektor dari parameter-parameter model. Masing-masing dari elemen $\boldsymbol{\eta}$ dapat dinyatakan dengan $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij},\ i = 1,2,\ldots,n$ dan $j = 1,2,\ldots,p$;
3. *link function* (fungsi penghubung) yaitu komponen menghubungkan antara komponen random dan komponen sistematik. Misalkan $\mu_i = E(Y_i)$ dimana $i = 1,2,\ldots,n$. Model untuk menghubungkan μ_i dengan η_i adalah $g(.)$ sehingga $g(\mu_i) = \eta_i$. Fungsi $g(.)$ menghubungkan $E(Y_i)$ dengan variabel prediktor yaitu $g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij},\ i = 1,2,\ldots,n$ dan $j = 1,2,\ldots,p$.

2.2 Sebaran *Poisson*

Percobaan *Poisson* adalah percobaan yang menghasilkan variabel acak bernilai numerik, yaitu banyak sukses selama selang waktu tertentu atau dalam daerah tertentu. Panjang selang waktu bermacam-macam seperti satu menit, satu hari, satu bulan dan lain sebagainya. Banyak sukses dalam suatu percobaan *Poisson* disebut variabel acak *Poisson*. Menurut Walpole dan Myers (1986), fungsi peluang sebaran *Poisson* dapat dinyatakan sebagaimana persamaan (2.1).

$$P(Y_i = y_i) = \begin{cases} \dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \text{untuk } y_i \;\; i = 0,1,2,..., \\ 0, \text{ untuk } y_i \text{ lain} \end{cases} \tag{2.1}$$

Sebaran *Poisson* adalah suatu sebaran untuk peristiwa yang memiliki peluang kejadian kecil, dimana kejadian tersebut tergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit. Rata- rata dan ragam variabel acak *Poisson* sama dengan μ.

Untuk mengetahui apakah data yang diambil berasal dari populasi yang mengikuti sebaran *Poisson* atau tidak, dapat menggunakan uji *Kolmogorov Smirnov*. Menurut Daniel (1989), statistik uji yang digunakan dapat dinyatakan sebagaimana persamaan (2.2).

$$D_n = \sup\left|F_n(x) - F_0(x)\right| \tag{2.2}$$

di mana :

D_n : jarak maksimum antara fungsi peluang kumulatif data dengan fungsi peluang kumulatif *Poisson*

$F_n(x)$: fungsi peluang kumulatif data

F_0 (x) : fungsi peluang kumulatif sebaran *Poisson*

Hipotesis yang digunakan dalam pengujian *Kolmogorov Smirnov* adalah sebagai berikut.

H_0 : Data berasal dari populasi yang mengikuti sebaran *Poisson*

H_1 : Data bukan berasal dari populasi yang mengikuti sebaran *Poisson*

Apabila nilai statistik uji D_n lebih besar dari nilai statistik *Kolmogorov Smirnov*$_{(\alpha,n)}$ maka H_0 ditolak.

2.3 Regresi *Poisson*

Regresi *Poisson* secara umum digunakan untuk menganalisis *count data* (data hitung). Menurut Hinde dan Demetrio (2007), pada regresi *Poisson* terdapat asumsi Y~ *Poisson* (μ), hal ini berarti variabel respon diasumsikan menyebar *Poisson* dengan parameter μ. Model regresi *Poisson* didapatkan dari sebaran *Poisson* yang mendefinisikan parameter μ sebagai variabel kovariat, dengan y_i adalah pengamatan ke-*i* dari variabel respon. Regresi *Poisson* kemudian digunakan untuk memodelkan suatu peristiwa yang relatif jarang terjadi pada satuan unit tertentu. Secara umum, persamaan regresi *Poisson* dapat dinyatakan sebagaimana persamaan (2.3).

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij},\ i = 1,2,...,n \text{ dan } j = 1,2,...,p \qquad (2.3)$$

di mana :

p : jumlah variabel prediktor

n : jumlah pengamatan

$\hat{\beta}$: parameter model regresi *Poisson* yang diestimasi

2.3.1 Overdispersion

Analisis regresi *Poisson* adalah analisis regresi yang termasuk bagian dari *Generalized Linear Model* (GLM). Regresi *Poisson* digunakan untuk data dengan variabel respon yang mengikuti sebaran *Poisson* (Y~ *Poisson*). Asumsi penting pada analisis ini adalah ragam harus sama dengan rata-rata disebut *equidispersion*. Namun pada beberapa penelitian kondisi ini tidak terpenuhi, sering ditemukan *count data* (data hitung) yang memiliki ragam lebih besar dari rata-rata disebut dengan *overdispersion*. Namun apabila ditemukan kondisi pada analisis regresi *Poisson* dengan ragam lebih kecil dari rata-rata maka disebut dengan *underdispersion* (Hilbe, 2011).

Menurut Hinde dan Demetrio (2007), terdapat beberapa kemungkinan tidak terpenuhi *equidispersion* pada suatu pemodelan, antara lain adalah keragaman hasil pengamatan (keragaman antar individu sebagai komponen yang tidak dijelaskan oleh model), korelasi antar respon individu, terjadi *clustering* (pengelompokan) dalam populasi dan variabel teramati yang dihilangkan. Konsekuensi dari tidak terpenuhi *equidispersion* adalah regresi *Poisson* tidak sesuai untuk memodelkan data karena model yang terbentuk akan menghasilkan estimasi parameter yang bias. Selain itu, *overdispersion* juga mengakibatkan nilai *standart error* menjadi lebih kecil *(underestimates)* dari seharusnya, sehingga menghasilkan kesimpulan yang tidak sesuai.

Pemeriksaan *overdispersion* dapat dilakukan menggunakan nilai *Deviance*. Ragam dari sebaran *Poisson* sama dengan rata-rata ($\sigma^2=\mu$). *Overdispersion* dideteksi menggunakan nilai *Deviance* dibagi dengan derajat bebas yang mempunyai nilai lebih besar dari 1, sedangkan *underdispersion* dideteksi dengan nilai *Deviance* dibagi dengan derajat bebas yang mempunyai nilai kurang dari 1. Menurut Agresti (2002), nilai *Deviance* dapat dinyatakan sebagaimana persamaan (2.4).

$$D = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \tag{2.4}$$

di mana :

n : jumlah pengamatan

y_i : variabel respon ke-i dengan i=1,2,...,n

$\hat{\mu}_i$: rata-rata variabel respon y yang di pengaruhi oleh nilai variabel prediktor pada pengamatan ke-i

2.3.2 *Zero Inflation*

Nilai nol yang berlebihan pada variabel respon *(zero inflation)* sering ditemukan pada analisis regresi *Poisson* baik untuk data diskrit atau *count data*. Apabila nilai nol memiliki arti penting dalam penelitian maka data tersebut tidak dapat dihilangkan namun harus dimasukkan dalam proses analisis. Pada beberapa penelitian dapat dijumpai kondisi terlalu banyak nol pada variabel respon yang lebih dari 50 persen. Menurut Famoye dan Singh (2006), besarnya proporsi data yang bernilai nol dapat berakibat pada ketepatan dari inferensia. Selain itu, regresi *Poisson* juga menjadi tidak tepat lagi memodelkan data yang sebenarnya.

2.4 Multikolinieritas

Multikolinieritas menunjukkan terdapat hubungan di antara beberapa atau semua variabel yang menjelaskan model regresi. Terdapat dua jenis multikolinieritas yaitu multikolinieritas sempurna dan multikolinieritas tidak sempurna. Pada multikolinieritas sempurna terdapat hubungan linier di antara variabel prediktor di mana satu variabel prediktor adalah fungsi linier dari variabel prediktor yang lain, sedangkan multikolinieritas tidak sempurna terjadi apabila terdapat hubungan linier yang tidak sempurna antar variabel prediktor (Gujarati, 1991).

Menurut Hocking (1996), pendeteksian multikolinieritas dapat dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Untuk regresi dengan lebih dari 2 variabel, persamaan untuk mengitung nilai VIF dapat dinyatakan sebagaimana persamaan (2.5).

$$VIF_j = \frac{1}{1-R_j^2} \tag{2.5}$$

di mana :

R_j^2 : Koefisien determinasi dari *auxiliary regression*

Auxiliary regression adalah regresi dengan X_j sebagai variabel respon, dan X selainnya sebagai variabel prediktor. Nilai R_j^2 berkisar antara 0 sampai dengan 1 sehingga nilai VIF akan naik seiring dengan kenaikan koefisien determinasi dari *auxiliary regression.* Nilai VIF yang lebih dari 5 merupakan bukti cukup untuk mendeteksi adanya multikolinieritas (Hocking, 1996).

Multikolinieritas sempurna yang terjadi dalam analisis regresi menyebabkan koefisien regresi menjadi *undetermined* (tidak dapat diestimasi), sedangkan pada multikolinieritas tidak sempurna *standart error* cenderung semakin besar seiring dengan meningkatnya tingkat korelasi antar variabel walaupun estimasi parameter masih dapat dilakukan dan bersifat BLUE *(Best Linear Unbiased Estimators)* serta tetap efisien. Besarnya *standart error* pada multikolinieritas tidak sempurna dapat menyebabkan selang kepercayaan menjadi lebih lebar. Selain itu, multikolinieritas tidak sempurna juga menyebabkan tanda untuk koefisien regresi berkebalikan dengan teori yang ada (Gujarati, 1991).

2.5 Regresi *Negative Binomial*

Regresi *Negative Binomial* adalah salah satu model regresi yang merupakan terapan dari GLM. Menurut Greene (2008), sebagai penerapan dari GLM maka sebaran *Negative Binomial* memiliki ketiga komponen yaitu komponen random, komponen sistematik dan fungsi penghubung. Pada model regresi *Negative Binomial*, variabel respon y_i diasumsikan memiliki sebaran *Negative Binomial* yang dihasilkan dari sebaran campuran *Poisson Gamma*. Fungsi peluang model regresi *Negative Binomial* dapat dinyatakan sebagaimana persamaan (2.6).

$$P(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right) y_i!} \left(\frac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{\kappa}} \left(\frac{\boldsymbol{\kappa}\mu_i}{1+\boldsymbol{\kappa}\mu_i}\right)^{y_i} \text{ dengan } i = 0,1,2,....,\text{n} \qquad (2.6)$$

Pada saat $\boldsymbol{\kappa} \to 0$ maka sebaran *Negative Binomial* memiliki ragam $V[Y] \to \mu$. Sebaran *Negative Binomial* akan mendekati suatu sebaran *Poisson* yang mengasumsikan rata-rata dan ragam yang sama yaitu $E[Y] = V[Y] = \mu$. Dalam model regresi *Negative Binomial*, y_i adalah variabel yang berupa *count data*. Menurut Hilbe (2011), model regresi *negative binomial* pada umumnya menggunakan fungsi penghubung logaritma atau *log link* yaitu:

$$\ln \mu_i = \mathbf{X}_i^T \boldsymbol{\beta}$$

Model regresi *Negative Binomial* dapat menggunakan *log link* karena $\ln \mu_i$ dan $\mathbf{X}_i^T \boldsymbol{\beta}$ akan terdefinisi di dalam interval (0, ∞) dan interpretasi parameter regresi akan menjadi lebih mudah. Setelah diperoleh fungsi penghubung yang tepat, maka selanjutnya dapat dinyatakan model regresi *Negative Binomial* untuk memodelkan *count data* yaitu:

$$\ln[E(Y_\text{i} \mid X_\text{i})] = \ln(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta} \text{ untuk i=1,2,…,n}$$

sehingga dapat diperoleh:

$$\mu_i = \exp(\mathbf{X}_i^T \boldsymbol{\beta})$$

2.6 Regresi *Zero Inflated Negative Binomial* (ZINB)

Model regresi *Zero Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari sebaran campuran *Poisson Gamma*. Menurut Garay *et al.* (2011), model ini dapat digunakan untuk memodelkan *count data* atau data diskrit dengan banyak nilai nol pada variabel respon (*zero inflation*) dan terjadi *overdispersion*. Jika y_i adalah variabel acak dengan i= 1,2,...*n* maka nilai dari variabel respon tersebut terjadi dalam dua keadaan. Keadaan pertama disebut *zero state* dan menghasilkan hanya pengamatan bernilai nol, sementara keadaan kedua disebut *negative binomial state* yang memiliki sebaran *Negative Binomial*. Fungsi peluang model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan sebagaimana persamaan (2.7).

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\dfrac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i = 0 \\ (1-\pi_i)\dfrac{\Gamma\left(y_i + \dfrac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\dfrac{1}{\boldsymbol{\kappa}}\right) y_i!}\left(\dfrac{1}{1+\boldsymbol{\kappa}\mu_i}\right)^{\frac{1}{\kappa}}\left(\dfrac{\boldsymbol{\kappa}\mu_i}{1+\boldsymbol{\kappa}\mu_i}\right)^{y_i}, \text{ untuk } y_i > 0 \end{cases} \tag{2.7}$$

dimana $0 \le \pi_i \le 1$, $\mu_i \ge 0$, $\boldsymbol{\kappa}$ adalah parameter dispersi dan $\Gamma(.)$ adalah fungsi *gamma*. Ketika $\pi_i = 0$, variabel acak y_i memiliki sebaran *Negative Binomial* dengan rata-rata μ_i dan parameter dispersi $\boldsymbol{\kappa}$, sehingga $Y_i \sim NB\ (\mu_i, \boldsymbol{\kappa})$. Diasumsikan bahwa μ_i dan π_i bergantung pada vektor dari variabel prediktor x_i yang dapat didefinisikan sebagai berikut.

$$\mu_i = e^{\mathbf{x}_i^T\boldsymbol{\beta}}$$

$$\pi_i = \frac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}, \text{ sehingga } (1\text{-}\pi_i) = \frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}$$

Model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dinyatakan sebagaimana persamaan (2.7) dan (2.8).

Model untuk *negative binomial state* $\hat{\mu}_i$

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}, i = 1,2,\ldots,n \text{ dan } j = 1,2,\ldots,p \tag{2.8}$$

Model untuk *zero inflation* $\hat{\pi}_i$

$$\text{logit } \hat{\pi}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij},\ i = 1,2,\ldots,n \text{ dan } j = 1,2,\ldots,p \tag{2.9}$$

di mana :

p : jumlah variabel prediktor

n : jumlah pengamatan

$\hat{\beta}$: parameter model regresi ZINB yang di estimasi

$\hat{\gamma}$: parameter model regresi ZINB yang di estimasi

Berdasarkan fungsi peluang untuk y_i yang telah diketahui pada persamaan (2.7), maka fungsi *likelihood* dan ln *likelihood* model regresi *Zero Inflated Negative Binomial* (ZINB) secara berurutan dapat dinyatakan sebagaimana persamaan (2.10) dan (2.11).

$$L(\kappa,\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \prod_{i=1}^{n} \dfrac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}} + \dfrac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\left(\dfrac{1}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i = 0 \\ \prod_{i=1}^{n} \dfrac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}} \dfrac{\Gamma\left(y_i + \dfrac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\dfrac{1}{\boldsymbol{\kappa}}\right) y_i!}\left(\dfrac{1}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\dfrac{\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}, \text{ untuk } y_i > 0 \end{cases} \tag{2.10}$$

$$\ln L(\kappa,\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{cases} \sum_{i=1}^{n} \ln\left(\dfrac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}} + \dfrac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\left(\dfrac{1}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i = 0\right) \\ \sum_{i=1}^{n} \ln\left(\dfrac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}} \dfrac{\Gamma\left(y_i + \dfrac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\dfrac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\dfrac{1}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\dfrac{\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa} e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}, \text{ untuk } y_i > 0\right) \end{cases}$$

$$
\begin{aligned}
&= \sum_{\substack{i=1 \\ y_i=0}}^{n} \ln\left(e^{\mathbf{x}_i^T \boldsymbol{\gamma}} + \left(\frac{1}{1+\kappa e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{\frac{1}{\kappa}} \right) - \sum_{i=1}^{n} \ln\left(1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}}\right) \\
&\quad + \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln\left(\Gamma\left(y_i + \frac{1}{\kappa} \right) \right) - \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln\left(\Gamma(y_i+1)\right) - \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln\left(\Gamma\left(\frac{1}{\kappa} \right) \right) \\
&\quad + \frac{1}{\kappa} \ln \sum_{\substack{i=1 \\ y_i>0}}^{n} \left(\frac{1}{1+\kappa e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + y_i \sum_{\substack{i=1 \\ y_i>0}}^{n} \ln\left(\frac{\kappa e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1+\kappa e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)
\end{aligned}
\tag{2.11}
$$

Estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan untuk memaksimalkan fungsi digunakan algoritma EM (*Expectation Maximization*). Fungsi peluang model regresi *Zero Infated Negative Binomial* (ZINB) terdiri dari dua kondisi yaitu $y_i=0$ dan $y_i>0$ dan telah diketahui bahwa variabel respon y_i pada penelitian ini juga terdiri dalam dua kondisi yaitu *zero state* dan *negative binomial state*. Untuk menggambarkan kondisi y_i secara terperinci, maka akan didefinisikan kembali variabel y_i dengan suatu variabel laten Z_i.

$$
Z_i = \begin{cases} 1 \text{ , jika } y_i \text{ berasal dari } \textit{zero state} \\ 0, \text{ jika } y_i > 0 \text{ berasal dari } \textit{negative binomial state} \end{cases}
\tag{2.12}
$$

Permasalahan pada pendefinisian ini adalah pada keadaan *negative binomial state*, Z_i dapat bernilai 0 atau 1. Permasalahan tersebut dapat diselesaikan menggunakan algoritma EM. Algoritma EM merupakan salah satu alternatif metode iteratif untuk memaksimumkan fungsi *likelihood* pada data yang mengandung variabel laten hasil pendefinisian variabel baru seperti variabel Z_i pada persamaan (2.12). Algoritma EM terdiri dari dua tahap yaitu tahap ekspektasi dan tahap maksimalisasi. Tahap ekspektasi yaitu tahap perhitungan ekspektasi dari fungsi ln *likelihood*, selanjutnya tahap maksimalisasi yaitu tahap perhitungan untuk mencari estimasi parameter yang memaksimumkan fungsi ln *likelihood* hasil dari tahap ekspektasi sebelumnya.

2.7 Pengujian Parameter Model

2.7.1 Pengujian Simultan

Pengujian parameter dilakukan untuk memeriksa peranan variabel prediktor dalam model. Parameter yang diuji pada pengujian simultan ini mencakup seluruh parameter **β** dan **γ** secara bersama-sama. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \beta_1 = \beta_2 = ... = \beta_p = \gamma_1 = \gamma_2 = ... = \gamma_p = 0$

H_1 : minimal ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0, j = 1, 2, ..., p$

Menurut Hosmer dan Lemeshow (2000), pengujian parameter model secara simultan dilakukan menggunakan statistik uji G. Statistik uji G adalah uji rasio kemungkinan maksimum (*likelihood ratio test*) yang digunakan untuk menguji peranan variabel prediktor di dalam model secara bersama-sama. Persamaan dari statistik uji G dapat dinyatakan sebagaimana persamaan (2.13).

$$G = -2\left[L_0 - L_p\right] \tag{2.13}$$

di mana :

L_o : ln *likelihood* model tanpa variabel prediktor (model intersep)

L_p : ln *likelihood* model dengan p variabel prediktor (model penuh)

Statistik uji G mengikuti sebaran χ^2 dengan derajat bebas p. Hipotesis nol ditolak apabila statistik uji G lebih besar dari $\chi^2_{p(\alpha)}$.

2.7.2 Pengujian Parsial

Pengujian koefisien regresi secara parsial digunakan untuk memeriksa pengaruh parameter regresi dari masing-masing variabel prediktor secara parsial pada model. Pengujian parsial parameter model meliputi pengujian parsial parameter **β** dan **γ.**

a. Pengujian Parsial Parameter **β**

Parameter yang diuji pada pengujian ini mencakup seluruh parameter **β** secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \beta_j = 0$
$H_1 : \beta_j \neq 0, j = 1, 2, ..., p$

Pengujian parameter regresi secara parsial dilakukan menggunakan statistik uji *t,* statistik uji ini sering digunakan untuk menguji signifikansi parameter regresi secara parsial pada masing-masing variabel prediktor. Menurut Hosmer dan Lemeshow (2000), persamaan dari statistik uji *t* dapat dinyatakan sebagaimana persamaan (2.14).

$$t_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{2.14}$$

Hipotesis nol ditolak jika nilai $|t|$ lebih besar atau sama dengan $t_{(\alpha/2;n-1)}$. Dengan α adalah tingkat taraf nyata yang digunakan. Pada statistik uji *t* terdapat $se(\hat{\beta}_j)$ yaitu *standard error* dari estimasi parameter $\hat{\beta}_j$ yang disebut sebagai matriks varian kovarian dari $\hat{\beta}_j$ yang diperoleh dari minus invers dari matriks Hessian.

b. Pengujian Parsial Parameter γ

Parameter yang diuji pada pengujian ini mencakup seluruh parameter γ secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \gamma_j = 0$
$H_1 : \gamma_j \neq 0, j = 1, 2, ..., p$

Pengujian parameter regresi secara parsial dilakukan menggunakan statistik uji *t,* statistik uji ini sering digunakan untuk menguji signifikansi parameter regresi secara parsial pada masing-masing variabel prediktor. Menurut Hosmer dan Lemeshow (2000), persamaan dari statistik uji *t* dapat dinyatakan sebagaimana persamaan (2.15).

$$t_j = \frac{\hat{\gamma}_j}{se(\hat{\gamma}_j)} \tag{2.15}$$

Hipotesis nol ditolak jika nilai $|t|$ lebih besar atau sama dengan $t_{(\alpha/2;n-1)}$. Dengan α adalah tingkat taraf nyata yang digunakan. Hipotesis nol ditolak jika nilai $|t|$ lebih besar atau sama dengan $t_{(\alpha/2;n-1)}$. Dengan α adalah tingkat taraf nyata yang digunakan. Pada statistik uji *t* terdapat $se(\hat{\gamma}_j)$ yaitu *standard error* dari estimasi parameter $\hat{\gamma}_j$ yang disebut sebagai matriks varian kovarian dari $\hat{\gamma}_j$ yang diperoleh dari minus invers dari matriks Hessian.

2.8 Kebaikan Model

Menurut Akaike (1978), nilai AIC (*Akaike Information Criterion*) dapat digunakan untuk pemilihan model terbaik. Nilai AIC dihitung berdasarkan nilai *maximum likelihood* dan jumlah parameter pada model regresi yang terbentuk. Persamaan untuk menghitung nilai AIC dinyatakan sebagaimana persamaan (2.16).

$$\mathrm{AIC} = -2\ln\ (maximum\ likelihood) + 2(number\ of\ \ parameter) \qquad (2.16)$$

Model regresi terbaik adalah model regresi yang mempunyai nilai AIC terkecil.

2.9 Penyakit *Tetanus Neonatorum*

2.9.1 Definisi Penyakit *Tetanus Neonatorum*

Menurut Saifuddin *et al*. (2006), penyakit *Tetanus Neonatorum* (*TN*) adalah penyakit Tetanus yang terjadi pada neonatus yaitu bayi berumur kurang dari satu bulan. *Tetanus Neonatorum* (*TN*) adalah penyakit yang disebabkan oleh bakteri *Clostridium Tetani* yang dapat menyebabkan kematian. *Clostridium Tetani* adalah bakteri yang mengeluarkan racun atau toksin dan menyerang sistem saraf pusat. Bakteri tersebut masuk ke dalam tubuh bayi melalui pintu masuk satu-satunya yaitu tali pusat yang dapat terjadi pada saat pemotongan tali pusat ketika bayi lahir maupun pada saat perawatan sebelum terlepasnya tali pusat. Sejak bakteri masuk ke dalam tubuh bayi sampai mulai timbulnya gejala (masa inkubasi) pada penyakit *Tetanus Neonatorum* adalah 3 sampai dengan 28 hari dan rata-rata masa inkubasinya adalah 6 hari. Apabila masa inkubasi penyakit tersebut kurang dari 7 hari, pada umumnya penyakit tersebut lebih parah dan dapat menyebabkan kematian yang tinggi. Untuk tercapainya target Eliminasi *Tetanus Neonatorum* di Provinsi Jawa Timur, maka perlu diketahui faktor-faktor risiko yang dapat menyebabkan terjadinya penyakit *Tetanus Neonatorum*.

2.9.2 Faktor Risiko Penyakit *Tetanus Neonatorum*

Beberapa faktor risiko yang dapat menyebabkan terjadinya penyakit *Tetanus Neonatorum* antara lain adalah sebagai berikut.

a. Kunjungan Ibu Hamil

Menurut Widagdo (2011), kunjungan ibu hamil adalah pertemuan atau kontak antara ibu hamil dan petugas kesehatan untuk mendapatkan pemeriksaan antenatal. Pemeriksaan antenatal adalah pemeriksaan kehamilan yang dilakukan untuk memeriksa keadaan ibu hamil dan janin secara berkala yang diikuti dengan pemeriksaan terhadap penyimpangan yang ditemukan pada janin. Pemeriksaan kehamilan dilakukan paling sedikit satu kali pada periode kehamilan tiga bulan pertama, paling sedikit satu kali pada periode kehamilan tiga bulan kedua dan paling sedikit dua kali pada periode kehamilan tiga bulan ketiga. Tujuan dari pemeriksaan ini adalah untuk menjaga agar ibu hamil dapat melalui masa kehamilan, persalinan dan nifas dengan baik dan selamat serta menghasilkan bayi yang sehat. Pemeriksaan kehamilan dilakukan oleh tenaga terlatih dan terdidik dalam bidang kebidanan, yaitu bidan dan perawat yang sudah terlatih. Pemeriksaan antenatal meliputi penimbangan berat badan, pengukuran tekanan darah, pengukuran tinggi fundus uteri, pemberian imunisasi TT dan pemberian tablet tambah darah. Berdasarkan rangkaian pemeriksaan antenatal, pemberian imunisasi TT adalah hal yang paling penting dilakukan untuk mencegah infeksi *Tetanus Neonatorum*. Dengan melakukan kunjungan kepada petugas kesehatan, diharapkan dapat memberikan informasi sedini mungkin kepada ibu hamil mengenai pentingnya menjaga kesehatan janin sampai dengan proses persalinan dilakukan. Oleh karena itu kunjungan ibu hamil merupakan faktor penting yang dapat mencegah terjadinya penyakit *Tetanus Neonatorum*.

b. Imunisasi Tetanus Toksoid (TT) Pada Ibu Hamil

Menurut Saifuddin *et al*. (2006), kekebalan terhadap penyakit *Tetanus Neonatorum* pada bayi dapat diperoleh melalui imunisasi Tetanus Toksoid (TT) pada ibu hamil. Pemberian imunisasi TT pada ibu hamil dimaksudkan agar bayi yang dilahirkan sudah mempunyai kekebalan terhadap toksin Tetanus yang didapatkan secara pasif pada saat bayi masih berada dalam kandungan. TT akan merangsang pembentukan antibodi spesifik yang mempunyai peranan penting dalam perlindungan terhadap Tetanus. Ibu hamil yang mendapatkan imunisasi TT dalam tubuhnya akan membentuk antibodi Tetanus. Antibodi Tetanus masuk dan menyebar melalui aliran darah janin ke seluruh tubuh janin. Sehingga dapat mencegah terjadinya penyakit *Tetanus Neonatorum*.

Tetanus Toksoid (TT) adalah antigen yang sangat aman untuk wanita hamil dan tidak menimbulkan bahaya bagi janin. Imunisasi TT untuk ibu hamil pada umumnya diberikan 2 kali. Jarak imunisasi TT dosis pertama (TT1) dan imunisasi TT dosis kedua (TT2) minimal adalah 4 minggu. Jarak pemberian TT pertama dan kedua serta jarak antara TT kedua dengan saat kelahiran sangat menentukan kadar antobodi Tetanus dalam darah bayi. Semakin lama jarak antara pemberian TT pertama dan kedua serta pemberian TT kedua dengan kelahiran bayi maka kadar antibodi Tetanus dalam darah bayi akan semakin tinggi. Hal ini dikarenakan jarak yang panjang pada pemberian TT akan memberikan waktu yang cukup untuk menyebarkan antobodi Tetanus dalam jumlah yang cukup dari tubuh ibu hamil ke tubuh bayi yang dikandung. Oleh karena itu pemberian imunisasi TT pada ibu hamil dianggap efektif untuk mencegah terjadinya penyakit *Tetanus Neonatorum*.

c. Jenis Penolong Persalinan

Pertolongan persalinan selain oleh tenaga kesehatan berhubungan erat dengan faktor risiko terjadinya penyakit *Tetanus Neonatorum*. Persalinan yang ditolong oleh tenaga kesehatan dapat mencegah terjadinya penyakit *Tetanus Neonatorum*. Hal ini dikarenakan pertolongan persalinan oleh tenaga kesehatan dapat lebih aman, bersih dan terjamin dibandingkan dengan persalinan yang ditolong oleh dukun bayi baik yang terlatih maupun yang tidak terlatih. Pertolongan persalinan yang bersih meliputi bersih tangan penolong, bersih daerah perineum ibu, jalan lahir tidak tersentuh oleh sesuatu yang tidak bersih, bersih alas tempat melahirkan dan memotong tali pusat menggunakan alat yang bersih.

Pertolongan persalinan oleh tenaga kesehatan di Indonesia masih cukup rendah. Masih cukup banyak persalinan yang ditolong oleh dukun bayi baik yang terlatih maupun yang tidak terlatih. Untuk mencegah terjadinya penyakit *Tetanus Neonatorum*, pemotongan tali pusat pada saat proses persalinan harus dilakukan menggunakan alat-alat yang steril. Namun penggunaan alat sederhana untuk memotong tali pusat seperti bilah bambu atau gunting yang tidak disterilkan terlebih dahulu, masih cukup banyak digunakan di Indonesia terutama pada persalinan yang tidak ditolong oleh tenaga kesehatan. Pemotongan tali pusat menggunakan alat-alat tersebut mengandung risiko tinggi terhadap terjadinya penyakit *Tetanus Neonatorum*. Selain alat pemotong tali pusat, hal lain yang perlu diperhatikan untuk mencegah terjadinya penyakit *Tetanus Neonatorum* adalah perawatan tali pusat. Merawat tali pusat berarti menjaga agar luka tersebut tetap bersih dan tidak terkena kotoran bayi. Untuk merawat tali pusat tidak diperbolehkan membubuhkan atau mengoleskan ramuan, abu dapur dan lain sebagainya pada luka tali pusat karena dapat menyebabkan infeksi dan menyebabkan terjadinya penyakit *Tetanus Neonatorum*. Oleh karena itu jenis penolong persalinan merupakan faktor penting yang dapat mencegah terjadinya penyakit *Tetanus Neonatorum* (Ngastiyah, 2003).

d. Kunjungan Neonatus

Kunjungan neonatus adalah pelayanan kesehatan sesuai standart yang diberikan oleh tenaga kesehatan yang kompeten kepada neonatus yaitu bayi umur 0-28 hari baik di fasilitas kesehatan maupun melalui kunjungan rumah. Kunjungan neonatus paling tidak dilakukan sebanya 3 kali selama periode 0 sampai 28 hari setelah bayi dilahirkan. Kunjungan neonatus pertama (KN1) dilakukan dalam kurun waktu 6-48 jam setelah bayi lahir, kunjungan neonatus kedua (KN2) dilakukan pada kurun waktu hari ke 3 sampai dengan hari ke 7 setelah bayi lahir dan kunjungan neonatus ketiga (KN3) dilakukan pada kurun waktu hari ke 8 sampai dengan hari ke 28 setelah bayi lahir. Kunjungan neonatus bertujuan untuk meningkatkan akses bayi terhadap pelayanan kesehatan dasar, mengetahui sedini mungkin bila terdapat kelainan pada bayi sehingga cepat mendapat pertolongan dan pemeliharaan kesehatan serta pencegahan penyakit melalui pemantauan pertumbuhan dan imunisasi. Oleh karena itu kunjungan neonatus merupakan faktor penting yang dapat mencegah terjadinya penyakit *Tetanus Neonatorum* (Saifuddin *et al*., 2006).

BAB 3
METODE PENELITIAN

Pada bab metode penelitian ini dijelaskan beberapa hal yang terkait dengan proses penelitian yang dilakukan antara lain yaitu: Sumber Data, Variabel Penelitian dan Definisi Operasional serta Metode Analisis untuk mencapai tujuan penelitian.

3.1 Sumber Data

Pada penelitian ini data sekunder yang digunakan bersumber dari Profil Kesehatan Provinsi Jawa Timur tahun 2012 yang dipublikasikan oleh DINKES (2013). Unit pengamatan pada penelitian ini adalah 38 Kabupaten/Kota di Provinsi Jawa Timur yang meliputi 29 Kabupaten dan 9 Kota. Jumlah penderita penyakit *Tetanus Neonatorum* di Provinsi Jawa Timur tahun 2012 adalah 29 orang.

3.2 Variabel Penelitian dan Definisi Operasional

Variabel respon (Y) yang digunakan pada penelitian ini adalah jumlah kasus *Tetanus Neonatorum* di setiap Kabupaten/Kota di Provinsi Jawa Timur, sedangkan variabel prediktor (X) yang digunakan adalah sebanyak 4 variabel. Definisi operasional dari masing-masing variabel respon dan variabel prediktor disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Definisi Operasional
Jumlah kasus *Tetanus Neonatorum* (Y)	Jumlah kasus *Tetanus Neonatorum* di setiap Kabupaten/Kota di Provinsi Jawa Timur.
Persentase kunjungan ibu hamil K4 (X_1)	Persentase kunjungan ibu hamil K4 di setiap Kabupaten/Kota di Provinsi Jawa Timur yang diperoleh dari persamaan. $\left(\frac{\text{Jumlah kunjungan ibu hamil k4}}{\text{Jumlah ibu hamil}} \text{x100\%}\right)$
Persentase imunisasi Tetanus Toksoid (TT) pada ibu hamil (X_2)	Persentase Imunisasi TT pada ibu hamil di setiap Kabupaten/Kota di Provinsi Jawa Timur yang diperoleh dari persamaan. $\left(\frac{\text{Jumlah imunisasi TT pada ibu hamil}}{\text{Jumlah ibu hamil}} \text{x100\%}\right)$

Tabel 3.1 Variabel Penelitian (lanjutan)

Variabel	Definisi Operasional
Persentase ibu bersalin ditolong tenaga kesehatan (X_3)	Persentase ibu bersalin ditolong tenaga kesehatan di setiap Kabupaten/Kota di Provinsi Jawa Timur Timur yang diperoleh dari persamaan. $\left(\frac{\text{Jumlah ibu bersalin ditolong nakes}}{\text{Jumlah ibu bersalin}} \text{x}100\%\right)$
Persentase kunjungan neonatus (X_4)	Persentase kunjungan neonatus di setiap Kabupaten/Kota di Provinsi Jawa Timur yang diperoleh dari persamaan. $\left(\frac{\text{Jumlah kunjungan neonatus}}{\text{Jumlah neonatus}} \text{x}100\%\right)$

Struktur data yang digunakan pada penelitian ini ditunjukkan pada Tabel 3.2.

Tabel 3.2 Struktur Data Penelitian

Kabupaten/Kota	Y	X_1	X_2	X_3	X_4
1	y_1	$x_{1.1}$	$x_{2.1}$	$x_{3.1}$	$x_{4.1}$
2	y_2	$x_{1.2}$	$x_{2.2}$	$x_{3.2}$	$x_{4.2}$
3	y_3	$x_{1.3}$	$x_{2.3}$	$x_{3.3}$	$x_{4.3}$
⋮	⋮	⋮	⋮	⋮	⋮
38	y_{38}	$x_{1.38}$	$x_{2.38}$	$x_{3.38}$	$x_{4.38}$

3.3 Metode Analisis

Analisis data dilakukan dengan menggunakan bantuan *software* statistika yaitu SPSS dan SAS. Untuk mencapai tujuan pertama pada penelitian ini maka langkah-langkah yang dilakukan adalah sebagai berikut.

a. Membentuk fungsi *likelihood* model regresi *Zero Inflated Negative Binomial* (ZINB) berdasarkan fungsi peluang yang telah diketahui.
b. Membentuk fungsi ln *likelihood* berdasarkan fungsi *likelihood* model regresi ZINB yang telah diketahui.
c. Membentuk sebaran bersama antara variabel respon y_i dan variabel laten Z_i.
d. Membentuk fungsi *likelihood* baru berdasarkan sebaran bersama antara variabel respon y_i dan variabel laten Z_i.
e. Membentuk fungsi ln *likelihood* berdasarkan fungsi *likelihood* baru dari sebaran bersama antara variabel respon y_i dan variabel laten Z_i.

f. Mencari estimasi parameter $\boldsymbol{\kappa}$, $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ secara terpisah, dengan memaksimumkan fungsi $\ln L(\boldsymbol{\kappa},\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})$ dan $\ln L(\boldsymbol{\gamma} \mid \mathbf{y},\mathbf{Z}^{(m)})$ hasil pemisahan fungsi ln *likelihood* pada langkah sebelumnya.

g. Mendapatkan estimasi parameter $\boldsymbol{\kappa}$, $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ yaitu $\hat{\boldsymbol{\kappa}}, \hat{\boldsymbol{\beta}}$ dan $\hat{\boldsymbol{\gamma}}$ dengan menggunakan algoritma EM (Ekspektasi Maksimalisasi).

Untuk mencapai tujuan kedua pada penelitian ini maka langkah-langkah yang dilakukan adalah sebagai berikut.

a. Melakukan analisis deskriptif pada variabel penelitian.

b. Memeriksa sebaran variabel respon apakah mengikuti sebaran *Poisson* atau tidak mengikuti sebaran *Poisson* sesuai dengan persamaan (2.2).

c. Memeriksa proporsi nilai nol pada variabel respon.

d. Memeriksa *overdispersion* dilakukan menggunakan nilai *Deviance* sesuai dengan persamaan (2.4).

e. Mengaplikasikan model regresi *Zero Inflated Negative Binomial* (ZINB) pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur tahun 2012 dengan variabel prediktor adalah faktor-faktor yang dianggap berpengaruh terhadap kasus *Tetanus Neonatorum*.

f. Pengujian signifikansi parameter model regresi. Pengujian dilakukan secara simultan dan secara parsial. Statistik uji yang digunakan untuk uji simultan adalah statistik uji G sesuai dengan persamaan (2.13) dan untuk uji secara parsial digunakan statistik uji *t* sesuai dengan persamaan (2.14) dan (2.15).

g. Menginterpretasi model regresi *Zero Inflated Negative Binomial* (ZINB) yang terbentuk.

h. Menentukan tingkat kebaikan model regresi *Zero Inflated Negative Binomial* (ZINB) yang terbentuk.

(Halaman ini sengaja dikosongkan)

BAB 4
ANALISIS DAN PEMBAHASAN

Pada bab ini dibahas mengenai proses estimasi parameter pada model regresi *Zero Inflated Negative Binomial* (ZINB). Selanjutnya model regresi ZINB digunakan untuk memodelkan jumlah kasus *Tetanus Neonatorum* di Provinsi Jawa Timur pada tahun 2012 serta mengetahui faktor-faktor apa saja yang berpengaruh terhadap jumlah kasus *Tetanus Neonatorum* di Provinsi Jawa Timur tahun 2012.

4.1 Estimasi Parameter Regresi *Zero Inflated Negative Binomial* (ZINB)

Pada model regresi *Zero Inflated Negative Binomial* (ZINB) yang telah dijelaskan pada sub bab 2.6 sebelumnya, diketahui bahwa setiap pengamatan $y_1, y_2, ..., y_n$ yang berasal dari variabel respon, terjadi dalam dua kondisi yang dapat dinyatakan sebagaimana persamaan (4.1).

$$y_i \sim \begin{cases} 0 & \text{, dengan peluang } \pi_i \\ y_i^* & \text{, dengan peluang } (1\text{-}\pi_i) \end{cases} \quad (4.1)$$

Kondisi pertama pada variabel respon disebut sebagai *zero state* yang terjadi dengan peluang π_i dan menghasilkan hanya pengamatan bernilai nol, sementara kondisi kedua disebut y_i^* (*negative binomial state*) yang terjadi dengan peluang (1-π_i). Fungsi peluang model regresi *Zero Inflated Negative Binomial* (ZINB) untuk $y_i = 0$ dan $y_i > 0$ dinyatakan sebagaimana persamaan (4.2).

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\dfrac{1}{1+\kappa\mu_i}\right)^{\frac{1}{\kappa}} & \text{, untuk } y_i = 0 \\ (1-\pi_i)\dfrac{\Gamma\left(y_i + \dfrac{1}{\kappa}\right)}{\Gamma\left(\dfrac{1}{\kappa}\right) y_i!}\left(\dfrac{1}{1+\kappa\mu_i}\right)^{\frac{1}{\kappa}}\left(\dfrac{\kappa\mu_i}{1+\kappa\mu_i}\right)^{y_i} & \text{, untuk } y_i > 0 \end{cases} \quad (4.2)$$

Di mana μ_i dan π_i adalah parameter dengan *link function* ln dan logit sebagaimana persamaan (4.3) dan (4.4).

$$\begin{bmatrix} \ln \mu_1 \\ \ln \mu_2 \\ \vdots \\ \ln \mu_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{1.1} + ... + \beta_n x_{1.k} \\ \beta_0 + \beta_1 x_{2.1} + ... + \beta_n x_{2.k} \\ \vdots \\ \beta_0 + \beta_1 x_{n.1} + ... + \beta_n x_{n.k} \end{bmatrix} \tag{4.3}$$

$$\begin{bmatrix} \text{logit } \pi_1 \\ \text{logit } \pi_2 \\ \vdots \\ \text{logit } \pi_n \end{bmatrix} = \begin{bmatrix} \gamma_0 + \gamma_1 x_{1.1} + ... + \gamma_n x_{1.k} \\ \gamma_0 + \gamma_1 x_{2.1} + ... + \gamma_n x_{2.k} \\ \vdots \\ \gamma_0 + \gamma_1 x_{n.1} + ... + \gamma_n x_{n.k} \end{bmatrix} \tag{4.4}$$

Berdasarkan persamaan (4.3) dan (4.4) diperoleh:

$$\ln \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

$$\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \tag{4.5}$$

dan

$$\text{logit } (\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\mathbf{x}_i^T \boldsymbol{\gamma} = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\left(1-\pi_i\right) e^{\mathbf{x}_i^T \boldsymbol{\gamma}} = \pi_i$$

$$e^{\mathbf{x}_i^T \boldsymbol{\gamma}} - \pi_i e^{\mathbf{x}_i^T \boldsymbol{\gamma}} = \pi_i$$

$$\pi_i \left(1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}\right) = e^{\mathbf{x}_i^T \boldsymbol{\gamma}}$$

$$\pi_i = \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} \tag{4.6}$$

Berdasarkan persamaan (4.5) dan (4.6), maka fungsi peluang pada persamaan (4.2) dapat dituliskan sebagaimana persamaan (4.7).

$$P(Y_i = y_i) = \begin{cases} \dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} + \dfrac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} \left(\dfrac{1}{1 + \boldsymbol{\kappa} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i = 0 \\ \dfrac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} \dfrac{\Gamma\left(y_i + \dfrac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\dfrac{1}{\boldsymbol{\kappa}}\right) y_i!} \left(\dfrac{1}{1 + \boldsymbol{\kappa} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}} \left(\dfrac{\boldsymbol{\kappa} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \boldsymbol{\kappa} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}\right)^{y_i}, \text{ untuk } y_i > 0 \end{cases} \tag{4.7}$$

X adalah matriks berukuran n x (p+1) yang berisi variabel-variabel prediktor yang berhubungan dengan peluang pada *zero state* dan *negative binomial state*, sedangkan $\boldsymbol{\kappa}, \boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ adalah vektor berukuran (p+1) x1 dari parameter regresi yang akan diestimasi. Bentuk fungsi *likelihood* model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dituliskan sebagaimana persamaan (4.8).

$$L(\kappa,\boldsymbol{\beta},\boldsymbol{\gamma})=\begin{cases}\prod_{i=1}^{n}\frac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}+\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i=0\\ \prod_{i=1}^{n}\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)y_i!}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}, \text{ untuk } y_i>0\end{cases} \quad (4.8)$$

Langkah berikutnya setelah mendapatkan fungsi *likelihood* dari model regresi *Zero Inflated Negative Binomial* (ZINB) adalah membentuk fungsi ln *likelihood* berdasarkan persamaan (4.8). Bentuk fungsi ln *likelihood* model regresi *Zero Inflated Negative Binomial* (ZINB) dapat dituliskan sebagaimana persamaan (4.9).

$$\ln L(\kappa,\boldsymbol{\beta},\boldsymbol{\gamma})=\begin{cases}\sum_{i=1}^{n}\ln\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}+\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}, \text{ untuk } y_i=0\right)\\ \sum_{i=1}^{n}\ln\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}, \text{ untuk } y_i>0\right)\end{cases}$$

$$\begin{aligned}&=\sum_{\substack{i=1\\y_i=0}}^{n}\ln\left(e^{\mathbf{x}_i^T\boldsymbol{\gamma}}+\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}\right)-\sum_{i=1}^{n}\ln\left(1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)+\sum_{\substack{i=1\\y_i>0}}^{n}\ln\left(\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)\right)\\&-\sum_{\substack{i=1\\y_i>0}}^{n}\ln\left(\Gamma(y_i+1)\right)-\sum_{\substack{i=1\\y_i>0}}^{n}\ln\left(\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\right)+\frac{1}{\boldsymbol{\kappa}}\ln\sum_{\substack{i=1\\y_i>0}}^{n}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)+y_i\sum_{\substack{i=1\\y_i>0}}^{n}\ln\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)\end{aligned} \quad (4.9)$$

Fungsi peluang model regresi *Zero Infated Negative Binomial* (ZINB) terdiri dari dua kondisi yaitu $y_i=0$ dan $y_i>0$ dan telah diketahui bahwa variabel respon y_i pada penelitian ini juga terdiri dalam dua kondisi yaitu *zero state* dan *negative binomial state*. Oleh karena itu, untuk menggambarkan kondisi y_i secara terperinci, maka akan didefinisikan kembali variabel y_i dengan suatu variabel laten Z_i.

$$Z_i = \begin{cases} 1 \text{ , jika } y_i \text{ berasal dari } \textit{zero state} \\ 0, \text{ jika } y_i > 0 \text{ berasal dari } \textit{negative binomial state} \end{cases} \tag{4.10}$$

Algoritma EM merupakan salah satu alternatif metode iteratif untuk memaksimumkan fungsi *likelihood* pada data yang mengandung variabel laten hasil pendefinisian variabel baru seperti variabel Z_i pada persamaan (4.10). Algoritma EM terdiri dari dua tahap yaitu tahap ekspektasi dan tahap maksimalisasi. Tahap ekspektasi yaitu tahap perhitungan ekspektasi dari fungsi ln *likelihood*, selanjutnya tahap maksimalisasi yaitu tahap perhitungan untuk mencari estimasi parameter yang memaksimumkan fungsi ln *likelihood* hasil dari tahap ekspektasi sebelumnya. Sebelum masuk pada tahap ekspektasi, terlebih dahulu ditentukan sebaran dari variabel laten Z_i yaitu sebagaimana persamaan (4.11).

$$P(Z_i = z_i) = \begin{cases} \pi_i & \text{, jika } z_i = 1 \\ 1 - \pi_i & \text{, jika } z_i = 0 \end{cases} \tag{4.11}$$

Pada saat $Z_i = 1$ maka peluang untuk Z_i akan sama dengan peluang y_i pada kondisi *zero state* yaitu sebesar π_i, sedangkan pada saat $Z_i = 0$ maka peluang untuk Z_i akan sama dengan peluang y_i pada kondisi *negative binomial state* yaitu sebesar $(1-\pi_i)$. Selanjutnya dibentuk sebaran bersama antara y_i dan Z_i yaitu sebagaimana persamaan (4.12).

$$f(y, z \mid \pi, \mu) = (\pi_i)^{z_i} (1-\pi_i)^{(1-z_i)} \left(\frac{\Gamma\left(y_i + \frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)\Gamma(y_i+1)} \left(\frac{1}{1+\kappa\mu_i}\right)^{\frac{1}{\kappa}} \left(\frac{\kappa\mu_i}{1+\kappa\mu_i}\right)^{y_i} \right)^{(1-z_i)} \tag{4.12}$$

Kemudian mensubtitusikan persamaan (4.5) dan (4.6) pada persamaan (4.12) sehingga diperoleh:

$$f(y,z\,|\,\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma})=\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)^{1-z_i}\left(\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}\right)^{1-z_i}$$

$$=\left(e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)^{-z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)\left(\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}\right)^{1-z_i}$$

$$=\left(e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)\left(\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}\right)^{1-z_i} \quad (4.13)$$

Fungsi *likelihood* baru dari sebaran bersama antara y_i dan Z_i pada persamaan (4.13) yaitu sebagaimana persamaan (4.14).

$$L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}\,|\,\mathbf{y},\mathbf{z})=\prod_{i=1}^{n}\left(e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)\left(\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\kappa}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}\right)^{1-z_i} \quad (4.14)$$

Diketahui bahwa $\dfrac{\Gamma(y+c)}{\Gamma(c)}=c(c+1)(c+2).....(c+y-1)$, sehingga diperoleh:

$$\frac{\Gamma\left(y_i+\frac{1}{\boldsymbol{\kappa}}\right)}{\Gamma\left(\frac{1}{\boldsymbol{\kappa}}\right)}=\left(\frac{1}{\boldsymbol{\kappa}}\right)\left(\frac{1}{\boldsymbol{\kappa}}+1\right)\left(\frac{1}{\boldsymbol{\kappa}}+2\right).....\left(\frac{1}{\boldsymbol{\kappa}}+y_i-1\right)$$

$$=\left(\frac{1}{\boldsymbol{\kappa}}\right)\left(\frac{1}{\boldsymbol{\kappa}}\right)(1+\boldsymbol{\kappa}).\left(\frac{1}{\boldsymbol{\kappa}}\right)(1+2\boldsymbol{\kappa}).....\left(\frac{1}{\boldsymbol{\kappa}}\right)\left(1+(y_i-1)\boldsymbol{\kappa}\right)$$

$$=\prod_{b=0}^{y_i-1}\left(b+\boldsymbol{\kappa}^{-1}\right) \quad (4.15)$$

dengan mensubtitusikan persamaan (4.15) pada persamaan (4.14), maka fungsi *likelihood* $L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}\,|\,\mathbf{y},\mathbf{z})$ menjadi persamaan baru sebagaimana persamaan (4.16).

$$L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}\,|\,\mathbf{y},\mathbf{z})=\prod_{i=1}^{n}\left(e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)^{z_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}\right)\left(\frac{\prod_{b=0}^{y_i-1}(b+\boldsymbol{\kappa}^{-1})}{y_i!}\left(\frac{1}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{\frac{1}{\boldsymbol{\kappa}}}\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)^{y_i}\right)^{1-z_i} \quad (4.16)$$

Sehingga fungsi ln *likelihood* baru dapat dituliskan sebagaimana persamaan (4.17).

$$\begin{aligned}\ln L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma}\,|\,\mathbf{y},\mathbf{z}) &= \sum_{i=1}^{n} Z_i\left(\mathbf{x}_i^T\boldsymbol{\gamma}\right)-\ln\left(1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)\\ &+\sum_{i=1}^{n}(1-Z_i)\left(\ln\left(\sum_{b=0}^{y_i-1}(b+\boldsymbol{\kappa}^{-1})\right)-\ln(y_l!)-\boldsymbol{\kappa}^{-1}\ln(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}})+y_i\ln\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)\right)\\ &= \sum_{i=1}^{n} Z_i\left(\mathbf{x}_i^T\boldsymbol{\gamma}\right)-\ln\left(1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)\\ &+\sum_{i=1}^{n}(1-Z_i)\left(\ln\left(\sum_{b=0}^{y_i-1}(b+\boldsymbol{\kappa}^{-1})\right)-\ln(y_l!)+y_i\ln\left(\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)-(y_i+\boldsymbol{\kappa}^{-1})\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)\right)\end{aligned} \quad (4.17)$$

Fungsi ln *likelihood* pada persamaan (4.17) adalah fungsi *likelihood* lengkap atau disebut juga dengan *complete likelihood*. Selanjutnya fungsi ln *likelihood* tersebut akan dimaksimumkan untuk mendapatkan hasil estimasi parameter **κ, β** dan **γ** dengan menggunakan algoritma EM.

Tahap ekspektasi dan maksimalilasasi dari algoritma EM dilakukan dengan langkah-langkah sebagai berikut.

1. Tahap ekspektasi yaitu menentukan nilai ekspektasi dari variabel Z_i, yaitu:

$$E(Z_i\,|\,y_{i,}\,\boldsymbol{\beta}^{(m)},\boldsymbol{\gamma}^{(m)})=Z_i^{(m)}$$

$$\begin{aligned}Z_i^{(m)} &= P(Z_i=1\,|\,y_i,\boldsymbol{\beta}^{(m)},\boldsymbol{\gamma}^{(m)})\\ &=\begin{cases}P(Z_i=1\,|\,y_i=0,\boldsymbol{\beta}^{(m)},\boldsymbol{\gamma}^{(m)}) & ,y_i=0\\ 0 & ,y_i>0\end{cases}\end{aligned}$$

Selanjutnya akan dicari terlebih dahulu nilai ekspektasi Z_i ketika $y_i=0$ yaitu:

$$P(Z_i=1\,|\,y_i,\boldsymbol{\beta}^{(m)},\boldsymbol{\gamma}^{(m)})$$

$$=\frac{P(y_i=0\,|\,Z_i=1)P(Z_i=1)}{P(y_i=0\,|\,Z_i=1)P(Z_i=1)+P(y_i=0\,|\,Z_i=0)P(Z_i=0)}$$

$$=\frac{\pi_i^{(m)}}{\pi_i^{(m)}+(1-\pi_i^{(m)})e^{-\mu_i^{(m)}}}$$

Kemudian mensubtitusikan persamaan (4.5) dan (4.6) sehingga diperoleh:

$$P(Z_i = 1 | y_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)})$$

$$= \frac{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}}{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}} + \dfrac{1}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}} e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}(m)}}}$$

$$= \frac{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}}{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}} + \dfrac{e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}(m)}}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}}$$

$$= \frac{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}}{\dfrac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)} + e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}(m)}}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}}$$

$$= \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)} + e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}(m)}}}$$

$$= \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)}}{e^{\mathbf{x}_i^T \boldsymbol{\gamma}(m)} + e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}(m)}}} \; \frac{e^{-(\mathbf{x}_i^T \boldsymbol{\gamma})(m)}}{e^{-(\mathbf{x}_i^T \boldsymbol{\gamma})(m)}}$$

$$= \frac{1}{1+\exp(-e^{\mathbf{x}_i^T \boldsymbol{\beta}^{(m)}} - \mathbf{x}_i^T \boldsymbol{\gamma}^{(m)})}$$

Dengan demikian, hasil ekspektasi Z_i dapat dituliskan sebagaimana persamaan (4.18).

$$Z_i^{(m)} = \begin{cases} \dfrac{1}{1+\exp(-e^{\mathbf{x}_i^T \boldsymbol{\beta}^{(m)}} - \mathbf{x}_i^T \boldsymbol{\gamma}^{(m)})} & , y_i = 0 \\ 0 & , y_i > 0 \end{cases} \tag{4.18}$$

Hasil ekspektasi Z_i pada persamaan (4.18) kemudian disubtitusikan pada persamaan (4.17). Hasil subtitusi dinyatakan sebagaimana persamaan (4.19).

$$\ln L(\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\gamma} \mid \mathbf{y},\mathbf{z}) = \sum_{i=1}^{n} Z_i^{(m)}\left(\mathbf{x}_i^T\boldsymbol{\gamma}\right) - \ln\left(1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right)$$
$$+\sum_{i=1}^{n}(1-Z_i)^{(m)}\left(\ln\left(\sum_{b=0}^{y_i-1}(b+\boldsymbol{\kappa}^{-1})\right) - \ln(y_l!) + y_i\ln\left(\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right) - (y_i+\boldsymbol{\kappa}^{-1})\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)\right) \tag{4.19}$$

Proses estimasi parameter **κ, β** dan **γ** dilakukan secara terpisah, sehingga fungsi *likelihood* lengkap dapat dituliskan kembali sebagaimana persamaan (4.20) dan (4.21).

$$\ln L(\boldsymbol{\kappa},\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)}) = \sum_{i=1}^{n}(1-Z_i)^{(m)}\left(\ln\left(\sum_{b=0}^{y_i-1}(b+\boldsymbol{\kappa}^{-1})\right) - \ln(y_l!) + y_i\ln\left(\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right) - (y_i+\boldsymbol{\kappa}^{-1})\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)\right) \tag{4.20}$$

dan

$$\ln L(\boldsymbol{\gamma} \mid \mathbf{y},\mathbf{Z}^{(m)}) = \sum_{i=1}^{n} Z_i^{(m)}\left(\mathbf{x}_i^T\boldsymbol{\gamma}\right) - \sum_{i=1}^{n}\ln\left(1+e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\right) \tag{4.21}$$

2. Tahap maksimalisasi meliputi maksimalisasi untuk parameter **κ**, **β** dan **γ.** Untuk proses maksimalisasi parameter digunakan metode iteratif numerik yaitu metode *Newton Raphson*. Tahap maksimalisasi untuk parameter **κ** dan **β** dilakukan secara bersama sama dimana **κ** (kappa) adalah parameter dispersi pada model regresi *Zero Inflated Negative Binomial* (ZINB). Untuk mendapatkan estimasi parameter **κ** dan **β**, fungsi ln *likelihood* pada persamaan (4.20) diturunkan secara parsial terhadap masing-masing parameter **κ** dan **β** . Turunan pertama fungsi ln *likelihood* pada persamaan (4.20) terhadap (**κ**) dinyatakan sebagaimana persamaan (4.22).

$$\frac{\partial \ln L(\boldsymbol{\kappa},\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})}{\partial(\boldsymbol{\kappa})}$$
$$=\sum_{i=1}^{n}(1-Z_i)^{(m)}\left(\sum_{b=0}^{y_i-1}-\boldsymbol{\kappa}^{-2}\frac{1}{(b+\boldsymbol{\kappa}^{-1})} + y_i\frac{1}{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}e^{\mathbf{x}_i^T\boldsymbol{\beta}} - \left(-\frac{\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}{\boldsymbol{\kappa}^2} + \frac{\left(y_i+\boldsymbol{\kappa}^{-1}\right)e^{\mathbf{x}_i^T\boldsymbol{\beta}})}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}})}\right)\right) \tag{4.22}$$

Selanjutnya turunan kedua fungsi ln *likelihood* pada persamaan (4.20) terhadap (**κ**) dinyatakan sebagaimana persamaan (4.23).

$$\frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})}{\partial(\boldsymbol{\kappa})\partial(\boldsymbol{\kappa})}$$

$$=\sum_{i=1}^{n}(1-Z_i)^{(m)}\sum_{b=0}^{y_i-1}\boldsymbol{\kappa}^{-3}\frac{\left(2b+\boldsymbol{\kappa}^{-1}\right)}{\left(b+\boldsymbol{\kappa}^{-1}\right)^2}-\frac{y_i}{\boldsymbol{\kappa}^2}+2\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)\boldsymbol{\kappa}^2}\right)-\frac{2\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}{\boldsymbol{\kappa}^3}+\frac{\left(e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)^2\left(y_i+\boldsymbol{\kappa}^{-1}\right)}{\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)^2} \quad (4.23)$$

Turunan pertama fungsi ln *likelihood* pada persamaan (4.20) terhadap ($\boldsymbol{\beta}^T$) dapat dinyatakan sebagaimana persamaan (4.24).

$$\frac{\partial \ln L(\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})}{\partial(\boldsymbol{\beta}^T)}$$

$$=\sum_{i=1}^{n}(1-Z_i^{(m)})\left(y_i\mathbf{x}_i^T - (y_i+\boldsymbol{\kappa}^{-1})\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\right)\mathbf{x}_i^T\right) \quad (4.24)$$

Selanjutnya turunan kedua fungsi ln *likelihood* pada persamaan (4.20) terhadap ($\boldsymbol{\beta}$) dinyatakan sebagaimana persamaan (4.25).

$$\frac{\partial^2 \ln L(\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})}{\partial(\boldsymbol{\beta}^T)\partial(\boldsymbol{\beta})}$$

$$=-\sum_{i=1}^{n}(1-Z_i^{(m)})\left((y_i+\boldsymbol{\kappa}^{-1})\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}})^2}\mathbf{x}_i^T\mathbf{x}_i\right) \quad (4.25)$$

Untuk turunan parsial kedua fungsi ln *likelihood* pada persamaan (4.20) terhadap parameter dispersi $\boldsymbol{\kappa}$ dan parameter regresi $\boldsymbol{\beta}$ dapat dinyatakan sebagaimana persamaan (4.26).

$$\frac{\partial^2 \ln L(\boldsymbol{\beta} \mid \mathbf{y},\mathbf{Z}^{(m)})}{\partial(\boldsymbol{\kappa})\partial(\boldsymbol{\beta})}$$

$$=\sum_{i=1}^{n}(1-Z_i)^{(m)}\sum_{b=0}^{y_i-1}\frac{\left(\frac{\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}}\mathbf{x}\right)}{\boldsymbol{\kappa}^2}-\frac{2\ln\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}{\boldsymbol{\kappa}^3}-\frac{\left(y_i+\boldsymbol{\kappa}^{-1}\right)\left(\mathbf{x}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}{\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}+\frac{\left(\boldsymbol{\kappa}\mathbf{x}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)\left(y_i+\boldsymbol{\kappa}^{-1}\right)e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{\left(1+\boldsymbol{\kappa}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)^2} \quad (4.26)$$

Metode iteratif numerik *Newton Raphson* digunakan untuk proses estimasi parameter koefisien regresi sehingga diperoleh solusi dari fungsi ln *likelihood.* Algoritma metode iteratif numerik *Newton Raphson* adalah sebagai berikut:

a. Menentukan taksiran awal untuk parameter $\hat{\boldsymbol{\beta}}^{(0)}$ yang diperoleh dengan metode OLS, yaitu :

$$\hat{\boldsymbol{\beta}}^{(0)} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

b. Membentuk vektor gradien ***g***,

$$\mathbf{g}^{T}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)_{(p+2)} = \left(\frac{\partial \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\kappa}} \frac{\partial \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathbf{0}}}, \frac{\partial \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathbf{1}}}, \ldots, \frac{\partial \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{p}}\right)_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(m)}}$$

dengan p adalah banyaknya parameter yang diestimasi.

c. Membentuk matriks Hessian H:

$$\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)_{(p+2)(p+2)} = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\kappa}^2} & \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\kappa}\partial \boldsymbol{\beta}_0} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\kappa}\partial \boldsymbol{\beta}_p} \\ & \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0^{\,2}} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0\partial \boldsymbol{\beta}_p} \\ & & \ddots & \vdots \\ \textit{simetris} & & & \frac{\partial^2 \ln L(\boldsymbol{\kappa},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p^{\,2}} \end{pmatrix}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(m)}}$$

d. Memasukkan nilai $\hat{\boldsymbol{\beta}}^{(0)}$ ke dalam elemen-elemen vektor **g** dan matriks **H** sehingga diperoleh vektor $\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(0)}\right)$ dan matriks $\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(0)}\right)$.

e. Melakukan iterasi mulai dari $m = 0$ sebagaimana pada persamaan berikut :

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} - \mathbf{H}^{-1}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)$$

f. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $|\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}| \leq \boldsymbol{\varepsilon}$, dimana $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumnya, misal 10^{-4}.

3. Sama halnya dengan tahap maksimalisasi parameter $\boldsymbol{\kappa}$ dan $\boldsymbol{\beta}$, untuk tahap maksimalisasi parameter $\boldsymbol{\gamma}$ juga digunakan metode iteratif numerik yaitu metode *Newton Raphson*. Untuk mendapatkan estimasi parameter $\boldsymbol{\gamma}$, fungsi ln *likelihood* pada persamaan (4.21) diturunkan secara parsial terhadap parameter $\boldsymbol{\gamma}$. Turunan pertama fungsi ln *likelihood* pada persamaan (4.21) terhadap ($\boldsymbol{\gamma}^T$) dinyatakan sebagaimana persamaan (4.27).

$$\frac{\partial \ln L(\boldsymbol{\gamma} \mid \mathbf{y}, \mathbf{Z}^{(m)})}{\partial(\boldsymbol{\gamma}^T)}$$

$$= \sum_{i=1}^{n} Z_i^{(m)} \mathbf{x}_i^T - \sum_{i=1}^{n} \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} \mathbf{x}_i^T \qquad (4.27)$$

Selanjutnya turunan kedua fungsi ln *likelihood* pada persamaan (4.21) terhadap ($\boldsymbol{\gamma}$) dinyatakan sebagaimana persamaan (4.28).

$$\frac{\partial^2 \ln L(\boldsymbol{\gamma} \mid \mathbf{y}, \mathbf{Z}^{(m)})}{\partial(\boldsymbol{\gamma}^T)\partial(\boldsymbol{\gamma})}$$

$$= -\sum_{i=1}^{n} \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}}}{\left(1 + e^{\mathbf{x}_i^T \boldsymbol{\gamma}}\right)^2} \mathbf{x}_i^T \mathbf{x}_i \qquad (4.28)$$

Metode iteratif numerik *Newton Raphson* digunakan untuk proses estimasi parameter koefisien regresi sehingga diperoleh solusi dari fungsi ln *likelihood.* Algoritma metode iteratif numerik *Newton Raphson* adalah sebagai berikut:

a. Menentukan taksiran awal untuk parameter $\hat{\boldsymbol{\gamma}}^{(0)}$ yang diperoleh dengan metode OLS, yaitu :

$$\hat{\boldsymbol{\gamma}}^{(0)} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

b. Membentuk vektor gradien ***g***,

$$\mathbf{g}^T\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)_{(p+1)} = \left(\frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0}, \frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1}, \cdots, \frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_p}\right)_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_{(m)}}$$

dengan p adalah banyaknya parameter yang diestimasi.

c. Membentuk matriks Hessian H:

$$\mathbf{H}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)_{(p+1)(p+1)} = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_0{}^2} & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_0 \partial \gamma_1} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_0 \partial \gamma_p} \\ & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_1{}^2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_1 \partial \gamma_p} \\ & & \ddots & \vdots \\ \textit{simetris} & & & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \gamma_p{}^2} \end{pmatrix}_{\boldsymbol{\gamma}=\boldsymbol{\gamma}(m)}$$

d. Memasukkan nilai $\hat{\boldsymbol{\gamma}}^{(0)}$ ke dalam elemen-elemen vektor **g** dan matriks **H** sehingga diperoleh vektor $\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)$ dan matriks $\mathbf{H}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)$.

e. Melakukan iterasi mulai dari $m = 0$ sebagaimana pada persamaan berikut :

$$\hat{\boldsymbol{\gamma}}^{(m+1)} = \hat{\boldsymbol{\gamma}}^{(m)} - \mathbf{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)$$

f. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $|\hat{\boldsymbol{\gamma}}^{(m+1)} - \hat{\boldsymbol{\gamma}}^{(m)}| \le \boldsymbol{\varepsilon}$, dimana $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumnya, misal 10^{-4}.

4.2 Aplikasi Regresi *Zero Inflated Negative Binomial* (ZINB)

4.2.1 Analisis Deskriptif Variabel Penelitian

Variabel respon yang digunakan pada penelitian ini adalah jumlah kasus *Tetanus Neonatorum* di Provinsi Jawa Timur tahun 2012. Jumlah kasus *Tetanus Neonatorum* paling banyak yaitu 7 kasus yang terjadi di Kabupaten Jember dan Kabupaten Bangkalan, sedangkan jumlah kasus *Tetanus Neonatorum* paling sedikit yaitu 0 kasus (tidak terjadi kasus *Tetanus Neonatorum)* yang terjadi di 29 Kabupaten/Kota antara lain adalah Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Lumajang, Kabupaten Banyuwangi, Kabupaten Pasuruan, Kabupaten Sidoarjo, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Lamongan, Kabupaten Gresik, Kabupaten Pamekasan, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota

Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya dan Kota Batu. Persentase nilai nol pada variabel respon memiliki persentase paling besar. Hal inilah yang menjadi fokus penelitian.

Pemodelan kasus *Tetanus Neonatorum* menggunakan model regresi ZINB menggunakan empat variabel prediktor yaitu persentase kunjungan ibu hamil K4 (X_1), persentase imunisasi Tetanus Toksoid (TT) pada ibu hamil (X_2), persentase ibu bersalin ditolong tenaga kesehatan (X_3) dan persentase kunjungan neonatus (X_4). Gambaran deskriptif dari masing-masing variabel prediktor disajikan pada Tabel 4.1.

Tabel 4.1 Analisis Deskriptif Variabel Prediktor

Variabel	Rata-rata	Stdev	Ragam	Minimum	Maximum
X_1	84,06	7,42	55,00	70,67	101,55
X_2	3,92	14,63	214,14	0	87,00
X_3	88,94	6,79	46,11	75,01	101,40
X_4	94,24	8,36	69,84	76,59	111,22

Berdasarkan Tabel 4.1 dapat diperoleh informasi bahwa pada tahun 2012 di Provinsi Jawa Timur, rata-rata persentase kunjungan ibu hamil K4 sebesar 84,06 persen dimana Kabupaten Lamongan memiliki persentase tertinggi dan Kabupaten Jember memiliki persentase terendah. Rata-rata persentase imunisasi Tetanus Toksoid sebesar 3,92 persen dimana Kabupaten Sumenep memiliki persentase tertinggi dan 21 Kabupaten/Kota memiliki persentase terendah antara lain yaitu Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Lumajang, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan, Kabupaten Sidoarjo, Kabupaten Mojokerto, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Lamongan, Kabupaten Gresik, Kabupaten Sampang, Kabupaten Pamekasan, Kota Blitar, Kota Malang dan Kota Pasuruan. Rata-rata persentase ibu bersalin ditolong tenaga kesehatan sebesar 88,94 persen dimana Kabupaten Lamongan memiliki persentase tertinggi dan Kota Kediri memiliki persentase terendah. Rata-rata persentase kunjungan neonatus sebesar 94,24 persen dimana Kabupaten Lamongan memiliki persentase tertinggi dan Kabupaten Jember memiliki persentase terendah. Berdasarkan nilai maksimum pada masing-masing variabel prediktor dapat diketahui bahwa terdapat persentase variabel prediktor yang memiliki nilai lebih dari 100 persen, menurut informasi dari Dinas Kesehatan Provinsi Jawa Timur hal ini dimungkinkan karena beberapa faktor salah satunya adalah lokasi fasilitas kesehatan yang terletak di daerah perbatasan sehingga memungkinkan penduduk dari wilayah Kabupaten/Kota lain melakukan pemeriksaan kesehatan bukan di daerah asal.

4.2.2 Pemeriksaan Sebaran Variabel Respon

Pemeriksaan sebaran variabel respon dilakukan untuk mengetahui variabel respon pada data mengikuti sebaran *Poisson* atau tidak. Menurut Daniel (1989), pemeriksaan sebaran variabel respon dilakukan menggunakan uji *Kolmogorov Smirnov*. Statistik uji yang digunakan adalah statistik uji D_n sebagaimana persamaan (2.2). Pemeriksaan sebaran variabel respon dilakukan pada 19 amatan pada variabel respon yang dipilih secara random dengan menghilangkan 19 amatan lain yang bernilai nol dan termasuk dalam keadaan *zero state*. Hasil pengujian *Kolmogorov Smirnov* secara lengkap dapat dilihat pada Lampiran 2A. Berdasarkan hasil pengujian diperoleh nilai statistik uji D_n sebesar 0,309 sedangkan nilai statistik uji *Kolmogorov Smirnov* $_{(0,05;19)}$ sebesar 0,312. Nilai statistik uji D_n (0,309) kurang dari statistik *Kolmogorov Smirnov* $_{(0,05;19)}$ (0,312), hal menunjukkan bahwa variabel respon mengikuti sebaran *Poisson*.

4.2.3 Pemeriksaan *Overdispersion*

Menurut Agresti (2002), pemeriksaan *overdispersion* regresi *Poisson* dilakukan menggunakan nilai *Deviance* dibagi dengan derajat bebas. Nilai *Deviance* dinyatakan sebagaimana persamaan (2.4). Kondisi *overdispersion* dideteksi menggunakan nilai *Deviance* dibagi dengan derajat bebas yang mempunyai nilai lebih besar dari 1. Hasil pemeriksaan *overdispersion* regresi *Poisson* secara lengkap dapat dilihat pada Lampiran 2B. Berdasarkan hasil pemeriksaan diperoleh nilai *Deviance* sebesar 51,124 dengan derajat bebas (db) sebesar 33, maka diperoleh nilai *Deviance* dibagi dengan derajat bebas sebesar 1,549. Nilai *Deviance* dibagi dengan derajat bebas memiliki nilai yang lebih besar dari 1, hal ini menunjukkan bahwa variabel respon mengalami *overdispersion*.

4.2.4 Pemeriksaan *Zero Inflation* Variabel Respon

Pemeriksaan *zero inflation* dilakukan dengan menghitung persentase amatan yang bernilai nol pada variabel respon. Hasil pemeriksaan *zero inflation* pada variabel respon disajikan pada Tabel 4.2.

Tabel 4.2 Hasil Pemeriksaan *Zero Inflation* pada Variabel Respon

Jumlah Kasus *Tetanus Neonatorum*	Frekuensi Jumlah Kasus *Tetanus Neonatorum*	Persentase	Kumulatif Persentase
0	29	76,3	76,3
1	3	7,9	84,2
2	2	5,3	89,5
3	1	2,6	92,1
5	1	2,6	94,7
7	2	5,3	100,0

Hasil pemeriksaan *zero inflation* variabel respon pada Tabel 4.2 menunjukkan bahwa terjadi *zero inflation* pada variabel respon karena persentase amatan bernilai nol lebih dari 50 persen yaitu sebesar 76,3 persen.

4.2.5 Pemeriksaan Multikolinieritas

Pemeriksaan multikolinieritas dilakukan untuk mengetahui hubungan diantara variabel prediktor yang menjelaskan model regresi. Menurut Hocking (1996), nilai yang digunakan sebagai acuan untuk pemeriksaan multikolinieritas adalah nilai VIF (*Variance Inflation Factor*). Persamaan untuk menghitung nilai VIF dinyatakan sebagaimana persamaan (2.5). Nilai VIF yang lebih dari 5 merupakan bukti cukup untuk mendeteksi multikolinieritas. Hasil pemeriksaan multikolinieritas disajikan pada Tabel 4.3. Untuk hasil pemeriksaan multikolinieritas secara lengkap dapat dilihat pada Lampiran 3.

Tabel 4.3 Hasil Pemeriksaan Multikolinieritas

Variabel Prediktor	Nilai VIF	Kesimpulan
X_1	3,156	Terdapat Multikolinieritas Antar Variabel Prediktor
X_2	1,034	
X_3	6,231	
X_4	3,962	

Hasil pemeriksaan multikolinieritas pada Tabel 4.3 menunjukkan bahwa terdapat multikolinieritas diantara variabel prediktor, karena pada variabel prediktor X_3 memiliki nilai VIF lebih dari 5. Namun, variabel prediktor X_3 tidak dihilangkan dari model karena diduga memberikan pengaruh signifikan terhadap jumlah kasus *Tetanus Neonatorum*.

4.2.6 Pemodelan Regresi *Negative Binomial* (NB)

Model regresi *Negative Binomial* (NB) adalah model regresi yang dapat digunakan untuk memodelkan data dengan variabel respon yang memiliki sebaran *Poisson* dan terjadi *overdispersion*. Sebelum mengaplikasikan model regresi ZINB pada kasus *Tetanus Neonatorum*, terlebih dahulu dilakukan pemodelan regresi *Negative Binomial* (NB). Pemodelan kasus *Tetanus Neonatorum* menggunakan model regresi *Negative Binomial* (NB) menggunakan empat variabel prediktor. Untuk mengetahui tingkat signifikansi hasil estimasi parameter pada model regresi *Negative Binomial* (NB), dilakukan pengujian signifikansi secara simultan dan parsial. Hasil estimasi parameter model regresi *Negative Binomial* (NB) pada kasus *Tetanus Neonatorum* serta nilai statistik uji G dan statistik uji *t* disajikan secara lengkap pada Tabel 4.4. Untuk hasil estimasi parameter secara lengkap dapat dilihat pada Lampiran 4.

Tabel 4.4 Hasil Estimasi Parameter Model Regresi *Negative Binomial* (NB)

Parameter	Estimasi	SE	*t* Hitung	(Pr >| *t*|)
$\hat{\beta}_0$	-11,798	5,879	-2,01	0,045
$\hat{\beta}_1$	-0,238	0,096	-2,48	0,013*
$\hat{\beta}_2$	0,021	0,020	1,04	0,299
$\hat{\beta}_3$	-0,011	0,155	-0,07	0,942
$\hat{\beta}_4$	0,330	0,132	2,50	0,012*
Statistik Uji G = 60,98				

*) Signifikan dengan taraf signifikansi 5 persen

Hasil pengujian signifikansi estimasi parameter model regresi *Negative Binomial* (NB) secara simultan dengan tingkat signifikansi sebesar 5 persen didasarkan pada statistik uji G. Nilai statistik uji G adalah 60,98. Nilai statistik uji G lebih besar dari $\chi^2_{(0,05;4)} = 9,488$. Hal ini menunjukkan bahwa secara simultan pada variabel prediktor X_1, X_2, X_3 dan X_4 memberikan pengaruh signifikan terhadap variabel respon. Sedangkan, hasil signifikansi estimasi parameter model regresi *Negative Binomial* (NB) secara parsial dengan tingkat signifikansi sebesar 5 persen didasarkan pada statistik uji *t*. Berdasarkan Tabel 4.4 terdapat dua variabel prediktor yang memiliki nilai *t* hitung yang lebih besar atau sama dengan $t_{(\alpha/2;37=2,00)}$ dan memiliki p-*value* kurang dari α (0,05). Hal ini menunjukkan bahwa variabel prediktor yang berpengaruh signifikan secara parsial pada model regresi *Negative Binomial* (NB) adalah persentase kunjungan ibu hamil (X_1) dan persentase kunjungan neonatus (X_4).

Persamaan model regresi *Negative Binomial* (NB) yang terbentuk adalah sebagai berikut.

$$\hat{\mu} = \exp(-11,798 - 0,238\ X_1 + 0,021\ X_2 - 0,011\ X_3 + 0,330\ X_4)$$

Interpretasi model regresi *Negative Binomial* (NB):

1. Setiap penambahan 1 persen kunjungan ibu hamil K4 (X_1) maka akan menurunkan rata-rata jumlah kasus *Tetanus Neonatorum* sebesar exp(0,238)=1,269 jika variabel lain bernilai konstan.
2. Setiap penambahan 1 persen imunisasi TT (Tetanus Toksoid) pada ibu hamil (X_2) maka akan meningkatkan rata-rata jumlah kasus *Tetanus Neonatorum* sebesar exp(0,021)=1,021 jika variabel lain bernilai konstan.
3. Setiap penambahan 1 persen ibu bersalin ditolong tenaga kesehatan (X_3) maka akan menurunkan rata-rata jumlah kasus *Tetanus Neonatorum* sebesar exp(0,011)=1,011 jika variabel lain bernilai konstan.
4. Setiap penambahan 1 persen kunjungan neonatus (X_4) maka akan meningkatkan rata-rata jumlah kasus *Tetanus Neonatorum* sebesar exp(0,330)=1,391 jika variabel lain bernilai konstan.

4.2.7 Pemodelan Regresi *Zero Inflated Negative Binomial* (ZINB)

Model regresi *Zero Inflated Negative Binomial* (ZINB) adalah model regresi yang dapat digunakan untuk memodelkan data dengan variabel respon yang memiliki sebaran *Poisson*, banyak amatan yang bernilai nol pada variabel respon dan terjadi *overdispersion*. Model ini merupakan pengembangan dari model regresi *Negative Binomial* (NB) untuk data dengan banyak amatan yang bernilai nol pada variabel respon (*zero inflation*). Model regresi ZINB diaplikasikan pada kasus *Tetanus Neonatorum* di Provinsi Jawa Timur. Pemodelan kasus *Tetanus Neonatorum* menggunakan model regresi ZINB menggunakan empat variabel prediktor yaitu persentase kunjungan ibu hamil K4 (X_1), persentase imunisasi Tetanus Toksoid (TT) pada ibu hamil (X_2), persentase ibu bersalin ditolong tenaga kesehatan (X_3) dan persentase kunjungan neonatus (X_4). Untuk mengetahui tingkat signifikansi hasil estimasi parameter pada model regresi ZINB, dilakukan pengujian signifikansi secara simultan dan parsial. Menurut Hosmer dan Lemeshow (2000), pengujian signifikansi hasil estimasi parameter pada model regresi ZINB secara simultan menggunakan statistik uji G dan pengujian signifikansi secara parsial menggunakan statistik uji *t*. Hasil estimasi parameter model regresi ZINB pada kasus *Tetanus Neonatorum* serta nilai statistik uji G dan statistik uji *t* disajikan secara lengkap pada Tabel 4.5. Untuk hasil estimasi parameter secara lengkap dapat dilihat pada Lampiran 5.

Tabel 4.5 Hasil Estimasi Parameter Model Regresi ZINB

Parameter	Estimasi	SE	*t* Hitung	(Pr >| *t*|)
$\hat{\beta}_0$	-5,847	3,602	-1,623	0,105
$\hat{\beta}_1$	-0,145	0,055	-2,644	0,008*
$\hat{\beta}_2$	-0,006	0,010	-0,599	0,549
$\hat{\beta}_3$	0,233	0,101	2,295	0,022*
$\hat{\beta}_4$	-0,023	0,067	-0,339	0,735
$\hat{\gamma}_0$	11,325	13,409	0,845	0,398
$\hat{\gamma}_1$	0,223	0,169	1,316	0,188
$\hat{\gamma}_2$	-0,296	0,179	-1,653	0,098
$\hat{\gamma}_3$	0,835	0,503	1,660	0,096
$\hat{\gamma}_4$	-1,078	0,539	-2,000	0,045*
Statistik Uji G = 581,24				

*) Signifikan dengan taraf signifikansi 5 persen

Hasil pengujian signifikansi estimasi parameter model regresi ZINB secara simultan dengan tingkat signifikansi sebesar 5 persen didasarkan pada statistik uji G. Nilai statistik uji G adalah 581,24. Nilai statistik uji G lebih besar dari $\chi^2_{(0,05;8)} = 15,507$. Hal ini menunjukkan bahwa secara simultan pada variabel prediktor X_1, X_2, X_3 dan X_4 memberikan pengaruh signifikan terhadap variabel respon. Untuk hasil signifikansi estimasi parameter model regresi ZINB secara parsial dengan tingkat signifikansi sebesar 5 persen didasarkan pada statistik uji *t*. Berdasarkan Tabel 4.5 terdapat dua variabel prediktor pada estimasi parameter model *negative binomial state* dan satu variabel prediktor pada estimasi parameter model *zero inflation* yang memiliki nilai *t* hitung yang lebih besar atau sama dengan $t_{(\alpha/2;37=2,00)}$ dan memiliki p-*value* kurang dari α (0,05). Hal ini menunjukkan bahwa variabel prediktor yang berpengaruh signifikan secara parsial pada model *negative binomial state* adalah persentase kunjungan ibu hamil (X_1) dan persentase ibu bersalin ditolong tenaga kesehatan (X_3), sedangkan variabel prediktor yang berpengaruh signifikan secara parsial pada model *zero inflation* adalah persentase kunjungan neonatus (X_4).

Persamaan model regresi *Zero Inflated Negative Binomial* (ZINB) yang terbentuk adalah sebagai berikut.

a. Model *negative binomial state* untuk $\hat{\mu}$

$$\hat{\mu} = \exp(-5,847 - 0,145\ X_1 - 0,006\ X_2 + 0,233\ X_3 - 0,023\ X_4)$$

b. Model *zero inflation* untuk $\hat{\pi}$

$$\hat{\pi} = \frac{\exp(11,325 + 0,223\ X_1 - 0,296\ X_2 + 0,835\ X_3 - 1,078\ X_4)}{1 + \exp(11,325 + 0,223\ X_1 - 0,296\ X_2 + 0,835\ X_3 - 1,078\ X_4)}$$

Interpretasi model *zero inflation* untuk $\hat{\pi}$ adalah sebagai berikut.

1. Setiap penambahan 1 persen kunjungan ibu hamil K4 (X_1) maka akan meningkatkan peluang jumlah kasus *Tetanus Neonatorum* sebesar exp(0,223)=1,249 kali dari jumlah kasus *Tetanus Neonatorum* semula, jika variabel lain bernilai konstan.
2. Setiap penambahan 1 persen imunisasi TT (Tetanus Toksoid) pada ibu hamil (X_2) maka akan menurunkan peluang jumlah kasus *Tetanus Neonatorum* sebesar exp(0,296)=1,344 kali dari jumlah kasus *Tetanus Neonatorum* semula, jika variabel lain bernilai konstan.
3. Setiap penambahan 1 persen ibu bersalin ditolong tenaga kesehatan (X_3) maka akan meningkatkan peluang jumlah kasus *Tetanus Neonatorum*

sebesar exp(0,835)=2,305 kali dari jumlah kasus *Tetanus Neonatorum* semula, jika variabel lain bernilai konstan.

4. Setiap penambahan 1 persen kunjungan neonatus (X_4) maka akan menurunkan peluang jumlah kasus *Tetanus Neonatorum* sebesar exp(1,078)=2,939 kali dari jumlah kasus *Tetanus Neonatorum* semula, jika variabel lain bernilai konstan.

Berdasarkan model *negative binomial state* dan *zero inflation* yang terbentuk, terdapat tanda dari koefisien regresi yang berkebalikan dengan teori yaitu persentase ibu bersalin ditolong tenaga kesehatan (X_3) untuk model *negative binomial state* dan persentase kunjungan ibu hamil K4 (X_1) serta persentase ibu bersalin ditolong tenaga kesehatan (X_3) untuk model *zero inflation*. Adanya koefisien regresi yang memiliki tanda berkebalikan dengan teori disebabkan oleh dampak adanya multikolinieritas. Selain itu tanda berkebalikan dengan teori juga disebabkan oleh bentuk pola data dari variabel prediktor tersebut yang memiliki korelasi positif dengan variabel respon. Bentuk pola data yang ditunjukkan dengan korelasi positif antara variabel respon dan variabel prediktor secara lengkap dapat dilihat pada Lampiran 6. Adanya multikolinieritas coba diatasi dengan Analisis Komponen Utama (AKU).

a. Mengatasi multikolinieritas menggunakan Analisis Komponen Utama (AKU)

Adanya multikolinieritas coba diatasi menggunakan Analisis Komponen Utama (AKU). Analisis Komponen Utama (AKU) adalah cara untuk mengelompokkan variabel-variabel prediktor yang korelasi liniernya sejalan linier menjadi satu komponen utama, sehingga dari p variabel prediktor akan didapat k komponen utama dimana $k \leq p$ yang dapat mewakili keragaman (variabilitas) variabel-variabel prediktor yang ada. Hasil estimasi parameter model regresi ZINB dengan menggunakan dua komponen utama yang terbentuk serta nilai statistik uji *t* disajikan secara lengkap pada Tabel 4.6. Untuk hasil analisis menggunakan Analisis Komponen Utama (AKU) secara lengkap dapat dilihat pada Lampiran 7 dan hasil estimasi parameter secara lengkap dapat dilihat pada Lampiran 8.

Tabel 4.6 Hasil Estimasi Parameter Model Regresi ZINB Menggunakan Komponen Utama yang Terbentuk.

Parameter	Estimasi	SE	*t* Hitung	(Pr >| *t*|)
$\hat{\beta}_0$	-0,431	4,781	-0,09	0,928
$\hat{\beta}_1$ (ku$_1$)	0,013	0,047	0,29	0,774
$\hat{\beta}_2$ (ku$_2$)	-0,002	0,016	-0,15	0,883
$\hat{\gamma}_0$	5,692	7,114	0,80	0,424
$\hat{\gamma}_1$ (ku$_1$)	0,028	0,144	0,20	0,845
$\hat{\gamma}_2$ (ku$_2$)	-1,068	1,778	-0,60	0,548

*) Signifikan dengan taraf signifikansi 5 persen

Persamaan model regresi *Zero Inflated Negative Binomial* (ZINB) menggunakan komponen utama yang terbentuk adalah sebagai berikut.

Model Regresi ZINB yang terbentuk adalah sebagai berikut :

a. Model *negative binomial* untuk $\hat{\mu}$

$$\hat{\mu} = \exp(-0,431 - 0,013\ \mathrm{ku}_1 - \ 0,002\ \mathrm{ku}_2)$$

$$\hat{\mu} = \exp(-0,431 - 0,013\ (0,348\ \mathrm{X}_1 - 0,025\ \mathrm{X}_2 + 0,371\ \mathrm{X}_3 + 0,355\ \mathrm{X}_4) \\ - \ 0,002\ (-0,058\ \mathrm{X}_1 + 0,992\ \mathrm{X}_2 + 0,073\ \mathrm{X}_3 + 0,051\ \mathrm{X}_4))$$

$$\hat{\mu} = \exp(-0,431 - 0,004\ \mathrm{X}_1 - \ 0,002\ \mathrm{X}_2 - \ 0,005\ \mathrm{X}_3 - 0,005\ \mathrm{X}_4)$$

b. Model *zero inflation* untuk $\hat{\pi}$

$$\hat{\pi} = \frac{\exp(5,692 + 0,028\ \mathrm{ku}_1 - 1,068\ \mathrm{ku}_2)}{1 + \exp(5,692 + 0,028\ \mathrm{ku}_1 - 1,068\ \mathrm{ku}_2)}$$

$$\hat{\pi} = \frac{\exp(5,692 + 0,028\ (0,348\ \mathrm{X}_1 - 0,025\ \mathrm{X}_2 + 0,371\ \mathrm{X}_3 + 0,355\ \mathrm{X}_4) - 1,068\ (-0,058\ \mathrm{X}_1 + 0,992\ \mathrm{X}_2 + 0,073\ \mathrm{X}_3 + 0,051\ \mathrm{X}_4))}{1 + \exp(5,692 + 0,028\ (0,348\ \mathrm{X}_1 - 0,025\ \mathrm{X}_2 + 0,371\ \mathrm{X}_3 + 0,355\ \mathrm{X}_4) - 1,068\ (-0,058\ \mathrm{X}_1 + 0,992\ \mathrm{X}_2 + 0,073\ \mathrm{X}_3 + 0,051\ \mathrm{X}_4))}$$

$$\hat{\pi} = \frac{\exp(5,692 + 0,072\ \mathrm{X}_1 - 1,060\ \mathrm{X}_2 - 0,068\ \mathrm{X}_3 - 0,045\ \mathrm{X}_4)}{1 + \exp(5,692 + 0,072\ \mathrm{X}_1 - 1,060\ \mathrm{X}_2 - 0,068\ \mathrm{X}_3 - 0,045\ \mathrm{X}_4)}$$

Berdasarkan pemodelan kasus *Tetanus Neonatorum* menggunakan model regresi ZINB dengan menggunakan dua komponen utama yang terbentuk diketahui bahwa masih terdapat tanda dari koefisien regresi yang berbeda dengan teori. Oleh karena itu pada penelitian ini diputuskan tidak mengunakan hasil pemodelan menggunakan komponen utama yang terbentuk.

4.2.8 Kebaikan Model

Menurut Akaike (1978), nilai AIC (*Akaike Information Criterion*) dapat digunakan untuk pemilihan model terbaik. Persamaan yang digunakan untuk menghitung nilai AIC dinyatakan sebagaimana persamaan (2.16). Untuk melihat kebaikan model, pada penelitian ini juga dilakukan pemodelan jumlah kasus *Tetanus Neonatorum* menggunakan model regresi *Negative Binomial* (NB) dan selanjutnya akan dilakukan perbandingan antara nilai AIC model regresi *Zero Inflated Negative Binomial* (ZINB) dan nilai AIC model regresi *Negative Binomial* (NB). Nilai AIC model regresi *Zero Inflated Negative Binomial* (ZINB) dan model regresi *Negative Binomial* (NB) disajikan pada Tabel 4.7.

Tabel 4.7. Nilai AIC Model Regresi ZINB dan Model Regresi NB

Model Regresi	Nilai AIC
Zero Inflated Negative Binomial (ZINB)	70,87
Negative Binomial (NB)	72,99

Nilai AIC untuk model regresi *Zero Inflated Negative Binomial* (ZINB) dan model regresi *Negative Binomial* (NB) pada Tabel 4.7 menunjukkan bahwa nilai AIC untuk model regresi Z*ero Inflated Negative Binomial* (ZINB) lebih kecil dibandingkan nilai AIC untuk model regresi *Negative Binomial* (NB). Sehingga dapat disimpulkan bahwa model regresi Z*ero Inflated Negative Binomial* (ZINB) lebih baik digunakan untuk memodelkan kasus *Tetanus Neonatorum* dibandingkan dengan model regresi *Negative Binomial* (NB).

DAFTAR PUSTAKA

Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.

Akaike, H. (1978). A Bayesian Analysis of The Minimum AIC Prosedure. *Annals of The Institute Statistical Mathematics*, Part A Page 9-14.

Daniel, W. W. (1989). *Statistik Non Parametrik Terapan* (A. T. K. W, Trans.). Jakarta: PT. Gramedia.

DINKES. (2013). Profil Kesehatan Provinsi Jawa Timur Tahun 2012. Surabaya: Dinas Kesehatan Provinsi Jawa Timur.

Famoye, F., & Singh, K. P. (2006). Zero Inflated Poisson Regression Model with an Applications Domestic Violence to Accident Data. *Journal of Data Science*, 117-130.

Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H. (2011). On Estimation and Influence Diagnostics for Zero Inflated Negative Binomial Regression Model. *Computational Statistics and Data Analysis, 55*, 1304-1318.

Greene, W. (2008). Functional Forms For The Negative Binomial Model For Count Data *Working Paper Department of Economics-Stren School of Business*, 585-590.

Gujarati, D. (1991). *Ekonometrika Dasar* (S. Zain, Trans.). Jakarta: Penerbit Erlangga.

Hilbe, J. M. (2011). *Negative Binomial Regression*. New York: Cambridge University Press.

Hinde, J., & Demetrio, C. G. B. (2007). *Overdispersi: Model and Estimation*. Exeter: Deparment of Laver Building.

Hocking, R. (1996). *Methods and Application of Linier Models*. New York: John Wiley and Sons.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons.

Lambert, D. (1992). Zero Inflated Poisson Regression, With an Application to Defect in Manufacturing. *Technometric, 34*(1).

Lestari, A. (2009). *Pemodelan Regresi Zero Inflated Poisson (Aplikasi Pada Data Pekerja Seks Komersial di Klinik Reproduksi Putat Jaya).* (Thesis), Institut Teknologi Sepuluh Nopember, Surabaya.

Lestari, S. P. (2014). *Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tetanus Neonatorum (TN) di Jawa Timur dengan Metode Zero Inflated Generalized Poisson Regression (ZIGP).* (Skripsi), Institut Teknologi Sepuluh Nopember, Surabaya.

Ngastiyah. (2003). *Perawatan Anak Sakit*. Jakarta: Penerbit Buku Kedokteran ECG.

Saifuddin, A. B., George, A., Gulardi, H. W., & Waspodo, D. (2006). *Pelayanan Kesehatan Maternal dan Neonatal*. Jakarta: Yayasan Bina Pustaka Sarwono Prawiroharjo.

Walpole, R. E., & Myers, R. H. (1986). *Ilmu peluang dan Statistika untuk Insinyur dan Ilmuwan* (R. K. Sembiring, Trans.). Bandung: Penerbit ITB.

Widagdo. (2011). *Masalah dan Tata Laksana Penyakit Infeksi Pada Anak*. Jakarta: CV. Agung Seto.

BAB 5
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan berdasarkan hasil penelitian ini adalah sebagai berikut.

1. Berdasarkan hasil penelitian, estimasi parameter model regresi *Zero Inflated Negative Binomial* (ZINB) dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan untuk memaksimalkan fungsi *likelihood* digunakan algoritma EM (*Expectation Maximization*).
2. Berdasarkan model regresi *Zero Inflated Negative Binomial* (ZINB) yang terbentuk variabel prediktor yang memberikan pengaruh signifikan terhadap jumlah kasus *Tetanus Neonatorum* meliputi persentase kunjungan ibu hamil K4 (X_1) dan persentase ibu bersalin ditolong tenaga kesehatan (X_3) untuk model *negative binomial state*, sedangkan untuk model *zero inflation* variabel prediktor yang memberikan pengaruh signifikan terhadap jumlah kasus *Tetanus Neonatorum* meliputi persentase kunjungan neonatus (X_4).

5.2 Saran

Berdasarkan hasil penelitian, saran yang bisa diberikan kepada Dinas Kesehatan Provinsi Jawa Timur adalah meningkatkan program K4 untuk ibu hamil, meningkatkan jumlah dan kualitas tenaga kesehatan dan meningkatkan program kunjungan neonatus untuk mengurangi jumlah kasus *Tetanus Neonatorum* serta memperbaiki kualitas kesehatan di Provinsi Jawa Timur.

Berdasarkan hasil penelitian, diketahui bahwa bentuk pola data menyebabkan tanda dari koefisien regresi berkebalikan dengan teori, oleh karena itu saran pada penelitian selanjutnya dapat menggunakan metode statistika yang memperhitungkan bentuk pola data seperti metode nonparametrik. Selain itu, dapat juga menggunakan model regresi lain untuk mengatasi masalah *overdispersion* dan *zero inflation* pada regresi *Poisson* seperti *Zero Inflated Poisson Invers Gaussian* (ZIPIG) atau dapat juga melakukan pemodelan dengan efek spasial dari tiap-tiap Kabupaten/Kota di Jawa Timur seperti metode *Geographically Weighted Zero Inflated Negative Binomial* (GWZINB).

(Halaman ini sengaja dikosongkan)

Lampiran 1. Data Penelitian Jumlah Kasus Tetanus Neonatorum di Provinsi Jawa Timur Tahun 2012

No.	Kabupaten/Kota	Y	X_1	X_2	X_3	X_4
1	Kabupaten Pacitan	0	90,01	0	92,59	96,71
2	Kabupaten Ponorogo	0	77,51	1,19	80,76	95,89
3	Kabupaten Trenggalek	0	83,64	0	98,88	103,93
4	Kabupaten Tulungagung	0	85,04	0	89,57	92,35
5	Kabupaten Blitar	0	84,42	0	89,26	94,32
6	Kabupaten Kediri	0	90,79	1,56	92,42	96,49
7	Kabupaten Malang	0	94,62	1,02	93,08	98,17
8	Kabupaten Lumajang	0	91,41	0	100,83	105,96
9	Kabupaten Jember	7	70,67	0	85,15	92,86
10	Kabupaten Banyuwangi	0	79,89	0	87,04	93,58
11	Kabupaten Bondowoso	1	91,61	0,36	90,80	103,73
12	Kabupaten Situbondo	3	75,21	0	82,08	90,53
13	Kabupaten Probolinggo	1	79,41	0	87,23	97,57
14	Kabupaten Pasuruan	0	82,8	0	86,02	91,78
15	Kabupaten Sidoarjo	0	80,87	0	84,94	82,17
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	Kabupaten Lamongan	0	101,55	0	101,40	108,92
25	Kabupaten Gresik	0	82,52	0	88,97	94,08
26	Kabupaten Bangkalan	7	93,98	10,55	98,98	108,55
27	Kabupaten Sampang	5	83,72	0	96,65	111,22
28	Kabupaten Pamekasan	0	90,3	0	90,74	98,36
29	Kabupaten Sumenep	2	76,05	87	86,95	91,48
30	Kota Kediri	0	75,15	0,63	75,01	76,59
31	Kota Blitar	0	73,53	0	82,46	78,18
32	Kota Malang	0	73,25	0	79,99	79,42
33	Kota Probolinggo	0	89,11	22,31	88,88	90,23
34	Kota Pasuruan	0	90,47	0	93,51	98,41
35	Kota Mojokerto	0	77,58	0,04	80,62	81,45
36	Kota Madiun	0	92,21	0,79	95,57	96,25
37	Kota Surabaya	0	84,69	1,01	81,24	85,05
38	Kota Batu	0	74,85	0,06	81,26	86,62

Lampiran 2A . Hasil Pemeriksaan Sebaran Variabel Respon *Software* SPSS

One-Sample Kolmogorov-Smirnov Test

		y
N		19
Poisson Parameter[a,b]	Mean	1,5263
Most Extreme Differences	Absolute	,309
	Positive	,309
	Negative	-,138
Kolmogorov-Smirnov Z		1,347
Asymp. Sig. (2-tailed)		,053

a. Test distribution is Poisson.

b. Calculated from data.

Lampiran 2B. Hasil Pemeriksaan *Overdispersion* Menggunakan *Software* SPSS

Goodness of Fit[a]

	Value	df	Value/df
Deviance	51,124	33	1,549
Scaled Deviance	51,124	33	
Pearson Chi-Square	65,046	33	1,971
Scaled Pearson Chi-Square	65,046	33	
Log Likelihood[b]	-38,220		
Akaike's Information Criterion (AIC)	86,439		
Finite Sample Corrected AIC (AICC)	88,314		
Bayesian Information Criterion (BIC)	94,627		
Consistent AIC (CAIC)	99,627		

Dependent Variable: Y

Model: (Intercept), x1, x2, x3, x4

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Lampiran 3. Hasil Pemeriksaan Multikolinieritas Menggunakan *Software* SPSS

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	-2,746	3,334		-,824	,416		
x1	-,167	,061	-,685	-2,757	,009	,317	3,156
x2	,013	,018	,106	,748	,460	,967	1,034
x3	,026	,093	,099	,282	,780	,160	6,231
x4	,161	,060	,744	2,672	,012	,252	3,962

a. Dependent Variable: Y

Lampiran 4. Hasil Estimasi Parameter Model Regresi NB Menggunakan *Software* SAS

```
Data datathesis;
Input y x1 x2 x3 x4;
Datalines;
0      90.01  0      92.59  96.71
0      77.51  1.19   80.76  95.89
0      83.64  0      98.88  103.93
0      85.04  0      89.57  92.35
0      84.42  0      89.26  94.32
0      90.79  1.56   92.42  96.49
0      94.62  1.02   93.08  98.17
0      91.41  0      100.83 105.96
7      70.67  0      85.15  92.86
0      79.89  0      87.04  93.58
1      91.61  0.36   90.80  103.73
3      75.21  0      82.08  90.53
1      79.41  0      87.23  97.57
0      82.8   0      86.02  91.78
0      80.87  0      84.94  82.17
0      78.89  0      86.57  91.09
0      86.56  2.37   90.33  95.93
0      84.46  0      93.32  94.96
0      73.31  0      75.06  87.99
0      82.04  1.25   85.52  89.85
0      92.26  0.48   93.92  99.07
2      92.45  0.74   98.40  102.6
1      87.47  17.76  93.76  98.74
0      101.55 0      101.40 108.92
0      82.52  0      88.97  94.08
7      93.98  10.55  98.98  108.55
5      83.72  0      96.65  111.22
0      90.3   0      90.74  98.36
2      76.05  87     86.95  91.48
0      75.15  0.63   75.01  76.59
0      73.53  0      82.46  78.18
0      73.25  0      79.99  79.42
0      89.11  22.31  88.88  90.23
0      90.47  0      93.51  98.41
0      77.58  0.04   80.62  81.45
0      92.21  0.79   95.57  96.25
0      84.69  1.01   81.24  85.05
0      74.85  0.06   81.26  86.62
;
run;
proc countreg data = datathesis type = negativebinom;
model y = x1 x2 x3 x4;
run;
```

Lampiran 4. Hasil Estimasi Parameter Model Regresi NB Menggunakan *Software* SAS (Lanjutan)

The COUNTREG Procedure

Model Fit Summary

Dependent Variable	y
Number of Observations	38
Log Likelihood	-30.49271
Maximum Absolute Gradient	5.26133E-7
Number of Iterations	19
AIC	72.98541
SBC	82.81093

Algorithm converged.

Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	-11.798226	5.879629	-2.01	0.0448
x1	-0.238485	0.096004	-2.48	0.0130
x2	0.021054	0.020299	1.04	0.2997
x3	-0.011202	0.154818	-0.07	0.9423
x4	0.330249	0.132147	2.50	0.0125
_Alpha	1.666733	1.015816	1.64	0.1008

Lampiran 5. Hasil Estimasi Parameter Model Regresi ZINB Menggunakan *Software* SAS

```
Data datathesis;
Input y x1 x2 x3 x4;
Datalines;
0      90.01  0      92.59  96.71
0      77.51  1.19   80.76  95.89
0      83.64  0      98.88  103.93
0      85.04  0      89.57  92.35
0      84.42  0      89.26  94.32
0      90.79  1.56   92.42  96.49
0      94.62  1.02   93.08  98.17
0      91.41  0      100.83 105.96
7      70.67  0      85.15  92.86
0      79.89  0      87.04  93.58
1      91.61  0.36   90.80  103.73
3      75.21  0      82.08  90.53
1      79.41  0      87.23  97.57
0      82.8   0      86.02  91.78
0      80.87  0      84.94  82.17
0      78.89  0      86.57  91.09
0      86.56  2.37   90.33  95.93
0      84.46  0      93.32  94.96
0      73.31  0      75.06  87.99
0      82.04  1.25   85.52  89.85
0      92.26  0.48   93.92  99.07
2      92.45  0.74   98.40  102.6
1      87.47  17.76  93.76  98.74
0      101.55 0      101.40 108.92
0      82.52  0      88.97  94.08
7      93.98  10.55  98.98  108.55
5      83.72  0      96.65  111.22
0      90.3   0      90.74  98.36
2      76.05  87     86.95  91.48
0      75.15  0.63   75.01  76.59
0      73.53  0      82.46  78.18
0      73.25  0      79.99  79.42
0      89.11  22.31  88.88  90.23
0      90.47  0      93.51  98.41
0      77.58  0.04   80.62  81.45
0      92.21  0.79   95.57  96.25
0      84.69  1.01   81.24  85.05
0      74.85  0.06   81.26  86.62
;
run;
proc countreg data = datathesis type = zinb;
model y = x1 x2 x3 x4 /
zi(link = logistic, var = x1 x2 x3 x4);
run;
```

Lampiran 5. Hasil Estimasi Parameter Model Regresi ZINB Menggunakan *Software* SAS (lanjutan)

The COUNTREG Procedure

Model Fit Summary

Dependent Variable	y
Number of Observations	38
Log Likelihood	-24.43536
Maximum Absolute Gradient	1.17282E-6
Number of Iterations	18
AIC	70.87072
SBC	88.88416

Algorithm converged.

Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	-5.847352	3.507019	-1.67	0.0954
x1	-0.145012	0.054447	-2.66	0.0077
x2	-0.006127	0.010210	-0.60	0.5485
x3	0.232751	0.101195	2.30	0.0214
x4	-0.022665	0.066548	-0.34	0.7334
Inf_Intercept	11.325414	13.389082	0.85	0.3976
Inf_x1	0.222748	0.169129	1.32	0.1878
Inf_x2	-0.295865	0.179011	-1.65	0.0984
Inf_x3	0.835209	0.502969	1.66	0.0968
Inf_x4	-1.077525	0.538635	-2.00	0.0454
_Alpha	0.085107	0.206312	0.41	0.6800

Lampiran 6. Hasil Korelasi Parsial antara Variabel Respon dan Variabel Prediktor

Correlations

		Y	x1	x2	x3	x4
Y	Pearson Correlation	1	-,092	,157	,180	,345*
	Sig. (2-tailed)		,584	,347	,280	,034
	N	38	38	38	38	38
x1	Pearson Correlation	-,092	1	-,097	,820**	,703**
	Sig. (2-tailed)	,584		,563	,000	,000
	N	38	38	38	38	38
x2	Pearson Correlation	,157	-,097	1	,004	-,022
	Sig. (2-tailed)	,347	,563		,982	,897
	N	38	38	38	38	38
x3	Pearson Correlation	,180	,820**	,004	1	,864**
	Sig. (2-tailed)	,280	,000	,982		,000
	N	38	38	38	38	38
x4	Pearson Correlation	,345*	,703**	-,022	,864**	1
	Sig. (2-tailed)	,034	,000	,897	,000	
	N	38	38	38	38	38

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Lampiran 7. Hasil Analisis Komponen Utama (AKU) Menggunakan *Software* SPSS

Communalities

	Initial	Extraction
x1	1,000	1,000
x2	1,000	1,000
x3	1,000	1,000
x4	1,000	1,000

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,596	64,904	64,904	2,596	64,904	64,904
2	1,005	25,121	90,024	1,005	25,121	90,024
3	,295	7,375	97,399	,295	7,375	97,399
4	,104	2,601	100,000	,104	2,601	100,000

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component			
	1	2	3	4
x1	,904	-,059	,409	,110
x2	-,065	,997	,045	,017
x3	,963	,073	-,044	-,256
x4	,920	,052	-,352	,161

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Component Score Coefficient Matrix

	Component			
	1	2	3	4
x1	,348	-,058	1,385	1,055
x2	-,025	,992	,154	,162
x3	,371	,073	-,149	-2,463
x4	,355	,051	-1,194	1,552

Lampiran 7. Hasil Analisis Komponen Utama (AKU) Menggunakan *Software* SPSS (lanjutan)

Component Score Covariance Matrix

Component	1	2	3	4
1	1,000	,000	,000	,000
2	,000	1,000	,000	,000
3	,000	,000	1,000	,000
4	,000	,000	,000	1,000

Extraction Method: Principal Component Analysis.
Component Scores.

Berdasarkan hasil analisis didapatkan persamaan komponen utama adalah sebagai berikut.

KU_1=0,348 X_1-0,025 X_2+0,371 X_3+0,355 X_4

KU_2 =-0,058 X_1+0,992 X_2+0,073 X_3+0,051 X_4

Lampiran 8. Hasil Estimasi Parameter Model Regresi ZINB dengan Komponen Utama yang Terbentuk Menggunakan *Software* SAS

```
Data datathesis;
Input y ku1 ku2 ;
Datalines;
0       100.01   6.47
0       90.95    7.47
0       102.69   7.67
0       95.61    6.32
0       95.98    6.43
0       100.10   7.95
0       102.29   7.33
0       106.83   7.46
7       89.15    6.85
0       93.31    6.49
1       102.38   6.96
3       88.76    6.25
1       94.63    6.74
0       93.31    6.16
0       88.83    5.70
0       91.91    6.39
0       97.63    8.82
0       97.72    6.76
0       84.60    5.71
0       92.14    7.31
0       102.11    7.03
2       105.08   7.79
1       99.83    24.42
0       111.63   7.07
0       95.12    6.51
7       107.70   17.78
5       104.47   7.87
0       100.01   6.40
2       89.02    92.91
0       81.15    5.65
0       83.94    5.74
0       83.36    5.64
0       95.46    28.05
0       101.11   6.60
0       85.82    5.58
0       101.69   7.32
0       89.78    6.36
0       86.94    6.07
;
run;
proc countreg data = datathesis type = zinb;
model y = ku1 ku2 /
zi(link = logistic, var = ku1 ku2);
run;
```

Lampiran 8. Hasil Estimasi Parameter Model Regresi ZINB dengan Dua Komponen Utama yang Terbentuk Menggunakan *Software* SAS (lanjutan)

```
The SAS System           05:32 Friday, May 12, 2015   2

                              The COUNTREG Procedure

                                Model Fit Summary

                    Dependent Variable                      y
                    Number of Observations                 38
                    Log Likelihood                  -35.19179
                    Maximum Absolute Gradient      1.17605E-6
                    Number of Iterations                   13
                    AIC                              84.38358
                    SBC                              95.84668


            Algorithm converged.


                              Parameter Estimates

                                          Standard                  Approx
            Parameter         Estimate       Error     t Value    Pr > |t|

            Intercept        -0.431072    4.781084       -0.09      0.9282
            ku1               0.013423    0.046775        0.29      0.7741
            ku2              -0.002391    0.016249       -0.15      0.8830
            Inf_Intercept     5.691563    7.113759        0.80      0.4237
            Inf_ku1           0.028276    0.144148        0.20      0.8445
            Inf_ku2          -1.067702    1.777877       -0.60      0.5481
            _Alpha            0.849296    1.101532        0.77      0.4407
```

Lampiran 9. Tahapan Algoritma EM Menggunakan Data Kecil pada *Negative Binomial State*

Berikut digunakan data kecil dengan satu variabel respon dan dua variabel prediktor yang akan diselesaikan menggunakan algoritma EM (*Expectation Maximization*).

Negative Binomial State

Y	X_1	X_2
0	2	4
2	4	5
1	5	3
0	3	6
3	5	3

a. Tahap Ekspektasi

Tahap ekspektasi adalah menentukan ekspektasi dari variabel laten *Zi* sebagaimana persamaan (4.18).

$$Z_i^{(m)} = \begin{cases} \dfrac{1}{1+\exp(-e^{\mathbf{x}_i^T\boldsymbol{\beta}^{(m)}} - \mathbf{x}_i^T\boldsymbol{\gamma}^{(m)})} & , y_i = 0 \\ 0 & , y_i > 0 \end{cases}$$

Berdasarkan persamaan diatas diperoleh nilai ekspektasi sebagai berikut.

$$Z_i^{(0)} = [4,481]$$

b. Tahap Maksimalisasi

Tahap maksimalisasi adalah memaksimumkan fungsi *likelihood* untuk mendapatkan estimasi parameter yang konvergen dengan menggunakan metode iteratif numerik *Newton Raphson*. Langkah-langkah yang dilakukan adalah sebagai berikut.

a. Menentukan taksiran awal untuk parameter $\hat{\boldsymbol{\gamma}}^{(0)}$ yang diperoleh dengan metode OLS, yaitu :

$$\hat{\boldsymbol{\beta}}^{(0)} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 10,87 & -1,31 & -1,35 \\ -1,31 & 0,21 & 0,12 \\ -1,35 & 0,12 & 0,21 \end{bmatrix}$$

Lampiran 9. Tahapan Algoritma EM Menggunakan Data Kecil pada *Negative Binomial State* (lanjutan)

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{bmatrix} 6 \\ 28 \\ 22 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}}^{(0)} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

$$= \begin{bmatrix} 10,87 & -1,31 & -1,35 \\ -1,31 & 0,21 & 0,12 \\ -1,35 & 0,12 & 0,21 \end{bmatrix} \begin{bmatrix} 6 \\ 28 \\ 22 \end{bmatrix} = \begin{bmatrix} -1,31 \\ 0,73 \\ -0,06 \end{bmatrix}$$

b. Membentuk matriks ***g*** sesuai dengan persamaan (4.24), sehingga diperoleh nilai sebagai berikut.

$$\mathbf{g}^{T}\left(\hat{\boldsymbol{\beta}}^{(m)}\right) = \left(\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0}, \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}, \cdots, \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p}\right)$$

$$\mathbf{g}^{T}\left(\hat{\boldsymbol{\beta}}^{(0)}\right) = \left[18,028 - 1,882\ 10,981\right]$$

c. Membentuk matriks Hessian H sesuai dengan persamaan (4.25) sehingga diperoleh nilai sebagai berikut.

$$\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)_{(p+1)(p+1)} = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0^{\,2}} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0 \partial \boldsymbol{\beta}_1} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0 \partial \boldsymbol{\beta}_p} \\ & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1^{\,2}} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_p} \\ & & \ddots & \vdots \\ \textit{simetris} & & & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p^{\,2}} \end{pmatrix}$$

$$\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(0)}\right) = \begin{pmatrix} 0,638 & 1,139 & 0,809 \\ 1,139 & -0,138 & -0,016 \\ 0,809 & -0,016 & 0,176 \end{pmatrix}$$

Lampiran 9. Tahapan Algoritma EM Menggunakan Data Kecil pada *Negative Binomial State* (lanjutan)

d. Membentuk matriks $\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)$ dan $\mathbf{H}^{-1}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)$

$$\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)=\begin{bmatrix}18,028\\-1,882\\10,981\end{bmatrix}$$

$$\mathbf{H}^{-1}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)=\begin{bmatrix}0,134 & 1,165 & -0,510\\1,165 & 2,960 & -5,087\\-0,510 & -5,087 & 7,564\end{bmatrix}$$

e. Melakukan iterasi mulai dari m = 0 sebagaimana pada persamaan berikut :

$$\hat{\boldsymbol{\beta}}^{(m+1)}=\hat{\boldsymbol{\beta}}^{(m)}-\mathbf{H}^{-1}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(m)}\right)$$

$$\hat{\boldsymbol{\beta}}^{(1)}=\hat{\boldsymbol{\beta}}^{(0)}-\mathbf{H}^{-1}\left(\hat{\boldsymbol{\beta}}^{(0)}\right)\mathbf{g}\left(\hat{\boldsymbol{\beta}}^{(0)}\right)$$

$$\hat{\boldsymbol{\beta}}^{(1)}=\begin{bmatrix}-1,31\\0,73\\-0,06\end{bmatrix}-\begin{bmatrix}0,134 & 1,165 & -0,510\\1,165 & 2,960 & -5,087\\-0,510 & -5,087 & 7,564\end{bmatrix}\begin{bmatrix}18,028\\-1,882\\10,981\end{bmatrix}$$

$$=\begin{bmatrix}-1,31\\0,73\\-0,06\end{bmatrix}-\begin{bmatrix}-5,37782\\-40,4226\\83,43689\end{bmatrix}$$

$$=\begin{bmatrix}4,07\\41,15\\-83,50\end{bmatrix}$$

f. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $|\hat{\boldsymbol{\beta}}^{(m+1)}-\hat{\boldsymbol{\beta}}^{(m)}|\leq\boldsymbol{\varepsilon}$, dimana $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumnya, misal 10^{-4}.

Lampiran 10. Tahapan Algoritma EM Menggunakan Data Kecil pada *Zero Inflation State*

Berikut digunakan data kecil dengan satu variabel respon dan dua variabel prediktor yang akan diselesaikan menggunakan algoritma EM (*Expectation Maximization*).

Zero Inflation State

Y	X_1	X_2
0	2	4
1	4	5
0	5	3
1	3	6
0	5	3

a. Tahap Ekspektasi

Tahap ekspektasi adalah menentukan ekspektasi dari variabel laten *Zi* sebagimana persamaan (4.18).

$$Z_i^{(m)} = \begin{cases} \dfrac{1}{1+\exp(-e^{\mathbf{x}_i^T \boldsymbol{\beta}^{(m)}} - \mathbf{x}_i^T \boldsymbol{\gamma}^{(m)})} & , y_i = 0 \\ 0 & , y_i > 0 \end{cases}$$

Berdasarkan persamaan diatas diperoleh nilai ekspektasi sebagai berikut.

$$Z_i^{(0)} = [2,962]$$

b. Tahap Maksimalisasi

Tahap maksimalisasi adalah memaksimumkan fungsi *likelihood* untuk mendapatkan estimasi parameter yang konvergen dengan menggunakan metode iteratif numerik *Newton Raphson*. Langkah-langkah yang dilakukan adalah sebagai berikut.

a. Menentukan taksiran awal untuk parameter $\hat{\boldsymbol{\gamma}}^{(0)}$ yang diperoleh dengan metode OLS, yaitu :

$$\hat{\boldsymbol{\gamma}}^{(0)} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 10,87 & -1,31 & -1,35 \\ -1,31 & 0,21 & 0,12 \\ -1,35 & 0,12 & 0,21 \end{bmatrix}$$

Lampiran 10. Tahapan Algoritma EM Menggunakan Data Kecil pada *Zero Inflation State* (lanjutan)

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{bmatrix} 12 \\ 47 \\ 57 \end{bmatrix}$$

$$\hat{\boldsymbol{\gamma}}^{(0)} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

$$= \begin{bmatrix} 10,87 & -1,31 & -1,35 \\ -1,31 & 0,21 & 0,12 \\ -1,35 & 0,12 & 0,21 \end{bmatrix} \begin{bmatrix} 2 \\ 7 \\ 11 \end{bmatrix} = \begin{bmatrix} -2,33 \\ 0,18 \\ 0,48 \end{bmatrix}$$

b. Membentuk matriks $\boldsymbol{g}$ sesuai dengan persamaan (4.27) sehingga diperoleh nilai sebagai berikut.

$$\mathbf{g}^{T}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right) = \left(\frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{\mathbf{0}}}, \frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{\mathbf{1}}}, \cdots, \frac{\partial \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{p}}\right)$$

$$\mathbf{g}^{T}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right) = \left[-4,44 - 0,329 - 3,544\right]$$

c. Membentuk matriks Hessian H sesuai dengan persamaan (4.28) sehingga diperoleh nilai sebagai berikut.

$$\mathbf{H}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)_{(p+1)(p+1)} = \begin{pmatrix} \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0^{\ 2}} & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0 \partial \boldsymbol{\gamma}_1} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0 \partial \boldsymbol{\gamma}_p} \\ & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1^{\ 2}} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\gamma}_p} \\ & & \ddots & \vdots \\ \mathit{simetris} & & & \frac{\partial^2 \ln L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_p^{\ 2}} \end{pmatrix}$$

$$\mathbf{H}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right) = \begin{pmatrix} -0,788 & 0,984 & -0,748 \\ 0,984 & 0,3 & 0,942 \\ -0,748 & 0,942 & 1,1 \end{pmatrix}$$

Lampiran 10. Tahapan Algoritma EM Menggunakan Data Kecil pada *Zero Inflation State* (lanjutan)

d. Membentuk matriks $\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)$ dan $\mathbf{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)$

$$\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)=\begin{bmatrix}-4,44\\-0,329\\-3,544\end{bmatrix}$$

$$\mathbf{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)=\begin{bmatrix}0,256 & 0,820 & -0,528\\0,820 & 0,654 & -0,003\\-0,528 & -0,003 & 0,552\end{bmatrix}$$

e. Melakukan iterasi mulai dari m = 0 sebagaimana pada persamaan berikut :

$$\hat{\boldsymbol{\gamma}}^{(m+1)}=\hat{\boldsymbol{\gamma}}^{(m)}-\mathbf{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(m)}\right)$$

$$\hat{\boldsymbol{\gamma}}^{(1)}=\hat{\boldsymbol{\gamma}}^{(0)}-\mathbf{H}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)$$

$$\hat{\boldsymbol{\gamma}}^{(1)}=\begin{bmatrix}-2,33\\0,18\\0,48\end{bmatrix}-\begin{bmatrix}0,256 & 0,820 & -0,528\\0,820 & 0,654 & -0,003\\-0,528 & -0,003 & 0,552\end{bmatrix}\begin{bmatrix}-4,44\\-0,329\\-3,544\end{bmatrix}$$

$$=\begin{bmatrix}-2,33\\0,18\\0,48\end{bmatrix}-\begin{bmatrix}0,466738\\-3,84395\\0,387383\end{bmatrix}$$

$$=\begin{bmatrix}-2,79\\4,03\\0,10\end{bmatrix}$$

f. Proses Iterasi akan berhenti jika telah diperoleh estimasi parameter yang konvergen dengan memenuhi $|\hat{\boldsymbol{\gamma}}^{(m+1)}-\hat{\boldsymbol{\gamma}}^{(m)}|\le\boldsymbol{\varepsilon}$, dimana $\boldsymbol{\varepsilon}$ adalah nilai yang sangat kecil dan telah ditetapkan sebelumnya, misal 10^{-4}.

DAFTAR LAMPIRAN

(Halaman ini sengaja dikosongkan)